BMC Systems Biology

BioMed Central
The Open Access Publisher

Constructing stochastic models from deterministic process equations by propensity adjustment

BMC Systems Biology 2011, **5**:187 doi:10.1186/1752-0509-5-187

Jialiang Wu (wjl@gatech.edu)
Brani Vidakovic (brani@bme.gatech.edu)
Eberhard O Voit (eberhard.voit@bme.gatech.edu)

ISSN 1752-0509

Article type Methodology article

Submission date 18 July 2011

Acceptance date 8 November 2011

Publication date 8 November 2011

Article URL http://www.biomedcentral.com/1752-0509/5/187

Constructing stochastic models from deterministic process equations by propensity adjustment

Jialiang Wu[1], Brani Vidakovic [2], Eberhard O Voit[2,3§]

[1]Deparment of Mathematics, Bioinformatics Program, Georgia Institute of Technology, Atlanta, GA30332, USA

[2] The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA30332, USA

[3]Integrative BioSystems Institute and The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA30332, USA

[§]Corresponding author

Email addresses:

JW: wjl@gatech.edu

BV: brani@bme.gatech.edu

EV: eberhard.voit@bme.gatech.edu

Original submission: July 2011

Revised submission: October 2011

Abstract

Background

Gillespie's stochastic simulation algorithm (SSA) for chemical reactions admits three kinds of elementary processes, namely, mass action reactions of 0^{th}, 1^{st} or 2^{nd} order. All other types of reaction processes, for instance those containing non-integer kinetic orders or following other types of kinetic laws, are assumed to be convertible to one of the three elementary kinds, so that SSA can validly be applied. However, the conversion to elementary reactions is often difficult, if not impossible. Within deterministic contexts, a strategy of model reduction is often used. Such a reduction simplifies the actual system of reactions by merging or approximating intermediate steps and omitting reactants such as transient complexes. It would be valuable to adopt a similar reduction strategy to stochastic modelling. Indeed, efforts have been devoted to manipulating the chemical master equation (CME) in order to achieve a proper propensity function for a reduced stochastic system. However, manipulations of CME are almost always complicated, and successes have been limited to relative simple cases.

Results

We propose a rather general strategy for converting a deterministic process model into a corresponding stochastic model and characterize the mathematical connections between the two. The deterministic framework is assumed to be a generalized mass action system and the stochastic analogue is in the format of the chemical master equation. The analysis identifies situations: where a direct conversion is valid; where internal noise affecting the system needs to be taken into account; and where the propensity function must be mathematically adjusted. The conversion from deterministic to stochastic models is illustrated with several representative examples, including reversible reactions with feedback controls, Michaelis-Menten enzyme kinetics, a genetic regulatory motif, and stochastic focusing.

Conclusions

The construction of a stochastic model for a biochemical network requires the utilization of information associated with an equation-based model. The conversion strategy proposed here guides a model design process that ensures a valid transition between deterministic and stochastic models.

Background

Most stochastic models of biochemical reactions are based on the fundamental assumption that no more than one reaction can occur at the exact same time. A consequence of this assumption is that only elementary chemical reactions can be converted directly into stochastic analogues [1]. These include: 1) zero-order reactions, such as the generation of molecules at a constant rate; 2) first-order reactions, with examples including elemental chemical reactions as well as transport and decay processes; and 3) second-order reactions, which include heterogeneous and homogeneous bimolecular reactions (dimerization). Reactions with integer kinetic orders other than 0, 1 and 2 are to be treated as combinations of sequential elementary reactions. The advantage of the premise of non-simultaneous reaction steps is that the stochastic reaction rate can be calculated from a deterministic, equation-based model with some degree of rigor, even though the derivation is usually not based on first physical principles but instead depends on other assumptions and on macroscopic information, such as a fixed rate constant in the equation-based model. The severe disadvantage is that this rigorous treatment is not practical for modelling larger biochemical reaction systems. The reasons include the following. First, in many cases, elementary reaction rates are not available. Secondly, even in the case that all reaction parameters are available, the computational expense is very significant when the system involves many species and reactions, and this fact ultimately leads to a combinatorial explosion of required computations. Within a

deterministic modelling framework, the common practice in this situation is to fit the transient and steady-state experimental data with a phenomenological, (differential) equation-based model, which explicitly or implicitly eliminates or merges some intermediate species and reactions. The best-known examples are probably Michaelis-Menten and Hill rate laws, which are ultimately explicit, but in truth approximate a multivariate system of underlying chemical processes.

Similar model reduction efforts have been carried out for stochastic modelling. For instance, the use of a complex-order function (which corresponds to a reduced equation-based model) was shown to be justified for some types of stochastic simulations. A prominent example is again the Michaelis-Menten rate law, which can be reduced from a system of elementary reactions to an explicit function by means of the *quasi-steady-state assumption* (see Result section and [2, 3]). However, model reduction within the stochastic framework has proven to be far more difficult than in the deterministic counterpart. The difficulties are mainly due to the fact that the reduction must be carried out on the chemical master equation (CME). This process is nontrivial and has succeeded only in simple cases.

In general, the construction of a stochastic model for a large biochemical network requires the use of information available from an equation-based model. In the past, several strategies have been proposed for this purpose and within the context of Gillespie's exact stochastic simulation algorithm (SSA; [1]) and its variants [4]. For example, Tian and Burrage [5] proposed that a stochastic model could be directly formulated from the deterministic model through a Poisson leaping procedure. However, a rigorous mathematical justification for such a conversion is lacking. Typical moment-

based approaches [6-8] derive ODEs for the statistical moments of the stochastic model from an equation-based model where the 0^{th}, 1^{st} and 2^{nd} order reactions follow mass action rate laws. More recently the moment method was extended to cover models consisting of rational rate laws [9]. Moreover, it was realized that the moment method is complementary to, but cannot fully replace, stochastic simulations, because it does not cover situations like genetic switches [6, 10].

In this article, we explore the mathematical connection between deterministic and stochastic frameworks for the pertinent case of Generalized Mass Action (GMA) systems, which are frequently used in Biochemical Systems Theory (BST; [11-13]). Specifically, we address two questions: First, under what conditions can a deterministic, equation-based model be converted directly into a stochastic simulation model? And second, what is a proper way of implementing this conversion? We will develop a method to answer these questions and demonstrate it for functions in the canonical power-law format of GMA systems. However, the results are applicable to other functions and formats as well, as we will demonstrate with several examples.

Representations of systems of biochemical reactions

Consider a well-stirred biochemical reaction system with constant volume and temperature, where N_s different chemical species $\{S_s\}_{s=1}^{N_s}$, interact through N_r unidirectional reaction channels $\{R_r\}_{r=1}^{N_r}$. Each reaction channel can be characterized as

$$R_r : \underline{\nu}_{r1}S_1 + \ldots + \underline{\nu}_{rN_s}S_{N_s} \xrightarrow{k_r} \bar{\nu}_{r1}S_1 + \ldots + \bar{\nu}_{rN_s}S_{N_s},$$

where $\underline{\nu}_{rs}$ and $\bar{\nu}_{rs}$ are the counts of molecular species S_s consumed and produced due to reaction R_r, respectively, and k_r is the rate constant. The changed amount of S_s

$v_{rs} = \bar{v}_{rs} - \underline{v}_{rs}$, which is due to the firing of reaction R_r, defines the stoichiometric coefficient of S_s in R_r. The stoichiometric coefficients of all species can be summarized according to each reaction R_r in the stoichiometric vector

$$\mathbf{v}_r \triangleq \begin{bmatrix} v_{r1} \\ \vdots \\ v_{rN_s} \end{bmatrix} \in \mathbb{Z}^{N_s}.$$

The stoichiometric vectors of all reactions can further be arranged as the stoichiometric matrix of the system

$$V \triangleq \left[\mathbf{v}_1, \ldots, \mathbf{v}_{N_r} \right] \in \mathbb{Z}^{N_s \times N_r}.$$

The size of the system is defined as $\Phi = AU$, where A is the Avogadro number and U is the reaction volume.

The modelling of biochemical reaction networks typically uses one of two conceptual frameworks: deterministic or stochastic. In a deterministic framework, the state of the system is given by the a non-negative vector $\left[\mathbf{X}(t)\right] = \left[\left[X_1(t)\right], \ldots, \left[X_{N_s}(t)\right] \right]^T \in \mathbb{R}^{N_s}$, where component $\left[X_s(t)\right]$ represents the concentration of species S_s, measured in moles per unit volume. The temporal evolution of the state of the system is modelled by a set of ordinary differential equations, which in our case are assumed to follow a generalized mass action (GMA) kinetic law. By contrast, in a stochastic framework, the state of the systems is characterized by a vector $\mathbf{x}(t) = \left[x_1(t), \ldots, x_{N_s}(t) \right]^T \in \mathbb{Z}^{N_s}$, whose values are non-negative integers. Specifically, $x_s(t) = \Phi\left[X_s(t)\right]$ is the count of S_s molecules, which is a

sample value of the random variable $X_s(t)$. The system dynamics of this process is typically described with the chemical master equation (CME). Both GMA and CME will be discussed in detail in the following sections.

Motivation for the power-law formalism: reactions in crowded media

Power-law functions with non-integer kinetics have proven very useful in biochemical systems analysis, and forty years of research have demonstrated their wide applicability (*e.g.*, see [11-13]). Generically, this type of description of a biochemical reaction can be seen either as a Taylor approximation in logarithmic space or as a heuristic or phenomenological model that has been applied successfully hundreds of times and in different contexts, even though it is difficult or impossible in many situations to trace it back to first mechanistic principles. A particularly interesting line of support for the power-law format can be seen in the example of a bimolecular reaction occurring in a spatially restricted environment. Savageau demonstrated that the kinetics of such a reaction can be validly formulated as a generalization of the law of mass action, where non-integer kinetic orders are allowed [14, 15]. Neff and colleagues [16-18] showed with careful experiments that this formulation is actually more accurate than alternative approaches.

Within the conceptual framework of power-law representations, the rate of the association reaction between molecules of species S_1 and S_2 is given as $k\left[X_1(t)\right]^{f_1}\left[X_2(t)\right]^{f_2}$. Here, *k* is the *rate constant* and f_1 and f_2 are real-valued *kinetic orders*, which are no longer necessarily positive integers as it is assumed in a mass action law. As an example, consider the reversible bimolecular reaction $S_1 + S_2 \underset{k_b}{\overset{k_f}{\rightleftharpoons}} S_3$. Like Neff

and colleagues [17], we begin by formulating a discrete update function for the population of S_3 molecules as

$$x_3(t+\Delta t) - x_3(t) = f\left([X_1],[X_2]\right)\Delta t\ x_1x_2\ -\ g\left([X_3]\right)\Delta t x_3. \qquad (1)$$

The first term on the right-hand side of this equation, $f\left([X_1],[X_2]\right)\Delta t\ x_1x_2$, describes the production of S_3: it depends on the totality of possible collisions x_1x_2 and also on some fraction $f\left([X_1],[X_2]\right)\Delta t$ that actually reacts and forms the product. In a dilute environment, $f\left([X_1],[X_2]\right)$ equals a traditional rate constant, and the reaction obeys the law of mass action, while in a spatially restricted environment, such as the cytoplasm, one needs to take crowding effects into account. As shown in Savageau [14, 15], the desired fraction of a reaction in a crowded environment becomes a *rate function* that depends on the current concentrations of S_1 and S_2. The second term, $g\left([X_3]\right)\Delta t x_3$, describes the fraction $g\left([X_3]\right)\Delta t$ of species S_3 that dissociates back into S_1 and S_2. This fraction may depend on some functional form of $[X_3]$ because in a crowded environment the complex may not be able to dissociate effectively. Thus, rate constants in the generalized mass action setting become rate functions (*cf.* [17]).

By taking the limit $\Delta t \to 0$, one obtains the differential equation

$$\frac{dx_3}{dt} = f\left([X_1],[X_2]\right)x_1x_2\ -\ g\left([X_3]\right)x_3. \qquad (2)$$

Savageau used Taylor series expansion to approximate the functions f and g in the logarithmic space $\left(\log[X_1],\log[X_2]\right)$ around some operating point (a,b). The result for f is

$$
\begin{aligned}
&\log f\left(\left[X_1\right],\left[X_2\right]\right) \triangleq F\left(\log\left[X_1\right],\log\left[X_2\right]\right) \\
&= F(a,b) + \left.\frac{\partial f\left(\left[X_1\right],\left[X_2\right]\right)}{\partial\left[X_1\right]}\right|_{(a,b)} (\log\left[X_1\right]-a) \\
&+ \left.\frac{\partial f\left(\left[X_1\right],\left[X_2\right]\right)}{\partial\left[X_2\right]}\right|_{(a,b)} (\log\left[X_2\right]-b) + \mathrm{HOT} \\
&\approx k_f + \alpha \log\left[X_1\right] + \beta \log\left[X_2\right],
\end{aligned}
\tag{3}
$$

where $k_f, \alpha,$ and β are constants related to the chosen operating point (*a*, *b*). The final step is achieved by ignoring all higher order terms (HOT) beyond the constant and linear terms. Transformation back to the Cartesian space yields

$$
f\left(\left[X_1\right],\left[X_2\right]\right) \approx k_a \left[X_1\right]^{\alpha} \left[X_2\right]^{\beta}, k_a = e^{k_f}. \tag{4}
$$

The same procedure leads to the power-law expression for the degradation term: $g\left(\left[X_3\right]\right) \approx k_d \left[X_3\right]^{\gamma}$. By combining constants we arrive at a power-law representation for the dynamics of species S_3 as

$$
\begin{aligned}
\frac{d\left[X_3\right]}{dt} &= k_a \left[X_1\right]^{\alpha} \left[X_2\right]^{\beta} \left[X_1\right]\left[X_2\right] - k_d \left[X_3\right]^{\gamma} \left[X_3\right] \\
&= k_a \left[X_1\right]^{a} \left[X_2\right]^{b} - k_d \left[X_3\right]^{c},
\end{aligned}
\tag{5}
$$

where $a = \alpha + 1, b = \beta + 1,$ and $c = \gamma + 1.$ As long as k_f, k_d, a, b and c remain more or less constant throughout a relevant range, the power-law model is mathematically well justified. In actual applications, the values of rate constants and kinetic orders can be estimated from experimental data [19]. When the functions *f* and *g* are originally not in power-law format, they can be locally approximated by power-law functions with a procedure similar to the one shown above (Equations (3) to (5)). An illustration will be given in the example section.

The Generalized Mass Action (GMA) format

In the GMA format within Biochemical Systems Theory, each process is represented as a univariate or multivariate power-law function. GMA models may be developed *de novo* or as an approximation of some other nonlinear rate laws. GMA models characterize the time evolution of the system state given that the system was in the state $X(t_0)$ at some initial time t_0. Generically, the state of the system is changed within a sufficiently small time interval by one out of the N_r possible reactions that can occur in the system. The reaction velocity through reaction channel R_r is:

$$\left[\frac{[X_1(t)]'}{v_{r1}} = \ldots = \frac{[X_{N_s}(t)]'}{v_{rN_s}}\right] = k_r \prod_{s=1}^{N_s} [X_s(t)]^{f_{rs}} \qquad (6)$$

for those $v_{rs} = \bar{v}_{rs} - \underline{v}_{rs} \neq 0, s = 1, \ldots, N_s$. As shown in the example of a bimolecular reaction, the kinetic order f_{rs} associated with species S_s captures the effects of both reactant properties (such as the stoichiometric coefficient $\underline{v}_{rs}$) and environmental influences (such as temperature, pressure, molecular crowding effects, etc.). Therefore f_{rs} does not necessary equal an integer $\underline{v}_{rs}$, which is assumed to be the case in mass action kinetics, but is possibly real-valued and may be negative. Summing up the contributions of all reactions, one obtains a GMA model describing the dynamics of S_s as

$$\frac{d}{dt}[X_s(t)] = \sum_{r=1}^{N_r} v_{rs} k_r \prod_{s=1}^{N_s} [X_s(t)]^{f_{rs}} \qquad (7)$$

for every $s = 1, \ldots, N_s$. Each reaction contributes either a production flux or a degradation flux to the dynamics of a certain species. Positive terms ($v_{rs} > 0$) represent the production of S_s, while negative terms ($v_{rs} < 0$) describe degradation. If f_{rs} is positive, then S_s accelerates the reaction R_r; a negative value represents that S_s inhibits the reaction, and

f_{rs}=0 implies that S_s has no influence on the reaction. The rate constant k_r for reaction R_r, is either positive or zero. Both, the rate constant and the kinetic order, are to be estimated from data.

Proper use of equation-based functions for stochastic simulations

The fundamental concept of a stochastic simulation is the propensity function $\alpha(\mathbf{X})$, and $\alpha(\mathbf{X})dt$ describes the probability that a reaction will change the value of a system variable within the next (infinitesimal) time interval $(t, t+dt)$. While a formal definition will be given later (Equation 18), it is easy to intuit that the propensity function is in some sense analogous to the rate in the corresponding deterministic model. In fact, the propensity function is traditionally assumed to be $\alpha(\mathbf{X}) = f_s(\mathbf{X})$, if the deterministic model is $X_s' = f_s(\mathbf{X}, t), s = 1, \ldots, N_s$. However, a proper justification for this common practice is by and large missing. Indeed, we will show that the direct use of a rate function as the propensity function in a stochastic simulation algorithm requires that at least one of the following assumptions be true:

1) f is a linear function;

2) the reaction is monomolecular;

3) all X_i in the system are noise-free variables, *i.e.*, without (or with ignorable) fluctuations, which implies that the covariance of any two participating reactants is zero (or close to zero).

Each of these assumptions constitutes a sufficient condition for the direct use of a rate function as the propensity function and applies, in principle, to GMA as well as other systems. The validity of these conditions will be discussed later. Specifically, the first

condition will be addressed in the Results section under the headings "0^{th}-order reaction kinetics" and "1^{st}-order reaction kinetics," while the second condition will be discussed under the heading "Real-valued order monomolecular reaction kinetics." The third condition will be the focus of Equations (29-36) and their associated explanations.

In reality, the rates of reactions in biochemical systems are commonly nonlinear functions of the reactant species, and fluctuations within each species are not necessarily ignorable. Therefore, to the valid use of an equation-based model in a stochastic simulation mandates that we know how to define a proper propensity function. The following section addresses this issue. It uses statistical techniques to characterize estimates for both the mean and variance of the propensity function, and these features will allow an assessment of the validity of the assumption $\alpha(\mathbf{X}) = f_s(\mathbf{X})$ and prescribe adjustments if the assumption is not valid.

Methods

Deriving the mean and variance of a power-law function of random variables

Consider a generic power-law function of random variables X_s with the format $PL(\mathbf{X}) = k\prod_{s=1}^{N_s} X_s^{f_s}$. Estimates of its mean μ_{PL} and variance σ_{PL} are given as

$$\mu_{PL} \approx k\prod_{s=1}^{N_s} \mu_s^{f_s} \exp\left(\sum_{i<j}^{N_s} f_i f_j \operatorname{cov}\left[\log X_i, \log X_j\right] \right) \tag{8}$$

$$\sigma_{PL}^{\,2} \approx \mu_{PL}^{\,2}\Omega \tag{9}$$

(for details, see Additional file 1). Here,

$$\Omega = \sum_{s=1}^{N_s} f_s \mu_s^{-2} \sigma_s^{\,2} + 2\sum_{i<j}^{N_s} f_i f_j \operatorname{cov}\left[\log X_i, \log X_j\right] \tag{10}$$

and μ_s=E$\left[X_s\right]$ and $\sigma_s^{\,2}=\mathrm{E}\left[(X_s-\mu_s)^2\right]$ are the mean and variance of random variable X_s, respectively. If we choose to express $\mathrm{cov}\left[\log X_i,\log X_j\right]$ as a function of μ_s, $\sigma_s^{\,2}$ and covariance $\sigma_{ij}=\mathrm{cov}\left[X_i,X_j\right]$, using a Taylor approximation, we obtain

$$\mu_{PL}\approx k\prod_{s=1}^{N_s}\mu_s^{\,f_s}\exp\left(-\frac{1}{2}\sum_{s=1}^{N_s}f_s\sigma_s^{\,2}/\mu_s^{\,2}+\frac{1}{2}\Omega\right) \tag{11}$$

$$\sigma_{PL}^{\,2}\approx\mu_{PL}^{\,2}\Omega, \tag{12}$$

where

$$\begin{aligned}\Omega\approx&\sum_{s=1}^{N_s}f_s\left(\sigma_s/\mu_s\right)^2+2\sum_{i<j}^{N_s}f_i f_j\left\{\sigma_{ij}/(\mu_i\mu_j)\right.\\&+\frac{1}{2}\log(\mu_i)\left(\sigma_j/\mu_j\right)^2+\frac{1}{2}\log(\mu_j)\left(\sigma_i/\mu_i\right)^2\\&\left.-\frac{1}{4}\left(\sigma_i/\mu_i\right)^2\left(\sigma_j/\mu_j\right)^2\right\}.\end{aligned} \tag{13}$$

Since many biochemical variables approximately follow a log-normal distribution [20-22], it is valuable to consider the special situation where $\left(X_1,\ldots,X_s\right)$ is log-normally distributed (*i.e.*, $\left(\log X_1,\ldots,\log X_s\right)$ is normally distributed). In such a case, a simpler alternative way to calculate $\mathrm{cov}\left[\log X_i,\log X_j\right]$ is

$$\mathrm{cov}\left[\log X_i,\log X_j\right]=\log\left(1+\frac{\sigma_{ij}}{\mu_i\mu_j}\right). \tag{14}$$

[23]. By substituting this result into (8)-(10), one obtains

$$\mu_{PL}\approx k\prod_{s=1}^{N_s}\mu_s^{\,f_s}\prod_{i<j}^{N_s}\left(1+\frac{\sigma_{ij}}{\mu_i\mu_j}\right)^{f_i f_j} \tag{15}$$

$$\sigma_{PL}^{\,2}\approx\mu_{PL}^{\,2}\Omega, \tag{16}$$

where

$$\Omega = \sum_{s=1}^{N_s} f_s \left(\frac{\sigma_s}{\mu_s} \right)^2 + 2 \sum_{i<j}^{N_s} f_i f_j \log \left(1 + \frac{\sigma_{ij}}{\mu_i \mu_j} \right). \qquad (17)$$

The approximation formulae for μ_{PL} and ${\sigma_{PL}}^2$ in eqns. (8)-(10) provide an easy numerical implementation if observation data are available to estimate $\mathrm{cov}\left[\log X_i, \log X_j \right]$. Furthermore, Equations (11)-(13) demonstrate how μ_{PL} and ${\sigma_{PL}}^2$ are related to μ_s, ${\sigma_s}^2$ and σ_{ij}; however, the price of this insight is paid by the possible inaccuracy introduced through the Taylor approximation. Equations (15)-(17) also provide a functional dependence of μ_{PL} and ${\sigma_{PL}}^2$ on $(\mu_s, {\sigma_s}^2, \sigma_{ij})$, but it is only valid if the additional assumption of log-normality is acceptable.

Deriving proper propensity functions for stochastic simulations from differential equation-based models

Assuming that the GMA model faithfully captures the average behaviour of a biochemical reaction system and recalling $\left[\mathbf{X}(t) \right] = \left([X_1(t)], \ldots, [X_{N_s}(t)] \right)^T$, the expected metabolite numbers are defined as the expectation

$$E\left[\mathbf{X} \right] = \left[\mathbf{X} \right] \Phi, \qquad (18)$$

where Φ is the system size as defined above.

To describe the reaction channel R_r stochastically, one needs the state update vector $\mathbf{v}_r$ and must characterize the quantity of molecules flowing through of reaction channel R_r during a small time interval. The key concept of this type of description is the propensity function $\alpha_r(\mathbf{x})$, which is defined as

$$\alpha_r(\mathbf{x})dt = \text{the probability that exactly one reaction } R_r \text{ will occur somewhere inside } U \text{ within infinitesmal interval } (t, t+dt), \text{ given current state } \mathbf{X}(t) = \mathbf{x}. \tag{19}$$

[1]. Because of the probabilistic nature of the propensity function, **X**(t) is no longer deterministic, and the result is instead stochastic and based on the transition probability

$$P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) = \text{Prob}\{\mathbf{X}(t) = \mathbf{x}, \text{ given } \mathbf{X}(t_0) = \mathbf{x}_0\}, \tag{20}$$

which follows the chemical master equation (CME)

$$\frac{\partial P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)}{\partial t} = \sum_{r=1}^{N_r} \left[\alpha_r(\mathbf{x} + \mathbf{v}_r) P(\mathbf{x} + \mathbf{v}_r, t \mid \mathbf{x}_0, t_0) - \alpha_r(\mathbf{x}) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \right] \tag{21}$$

Updating CME requires knowledge of every possible combination of all species counts within the population, which immediately implies that it can be solved analytically for only a few very simple systems and that numerical solutions are usually prohibitively expensive [24]. To address the inherent intractability of CME, Gillespie developed an algorithm, called the *Stochastic Simulation Algorithm* (SSA), to simulate CME models [1]. SSA is an exact procedure for numerically simulating the time evolution of a well-stirred reaction system. It is rigorously based on the same microphysical premise that underlies CME and gives a more realistic representation of a system's evolution than a deterministic reaction rate equation represented by ODEs. SSA requires knowledge of the propensity function, which however is truly available only for elementary reactions. These reactions include: 1) 0^{th} order reactions, exemplified with the generation of a molecule at a constant rate; 2) 1^{st} order monomolecular reactions, such as an elemental chemical conversion or decay of a single molecule; 3) 2^{nd} order bimolecular reactions, including reactive collisions between two molecules of the same or different species. The

reactive collision of more than two molecules at exactly the same time is considered highly unlikely and modelled as two or more sequential bimolecular reactions.

For elementary reactions, the propensity function of reaction R_r is computed as the product of a stochastic rate constant c_r and the number h_r of distinct combinations of reactant molecules, i.e.

$$\alpha_r(\mathbf{x}) = c_r h_r(\mathbf{x}), \quad r = 1,\ldots,N_r. \tag{22}$$

Here $h_r(\mathbf{x}) = \begin{cases} \prod_{s=1}^{N_s} \binom{x_s}{\underline{\nu}_{rs}} \approx \dfrac{\prod_{s=1}^{N_s} x_s^{\underline{\nu}_{rs}}}{\prod_{s=1}^{N_s} \underline{\nu}_{rs}!}, & \text{for } x_s \geq \underline{\nu}_{rs} > 0 \\ 0, \quad \text{otherwise} \end{cases}$, where x_s is the sample value of random variable X_s. The approximation is invoked when x_s is large and $(x_s - 1),\ldots,(x_s - \underline{\nu}_{rs} + 1)$ are approximately equal to x_s.

In Gillespie's original formulation [1] c_r is a constant that only depends on the physical properties of the reactant molecules and the temperature of the system, and $c_r dt$ is the probability that a particular combination of reactant molecules will react within the next infinitesimally small time interval $(t, t+dt)$. The constant c_r can be calculated from the corresponding deterministic rate constants, if they are known.

Since the assumption of mass action kinetics is not valid generally, especially in spatially restricted environments and in situations dominated by macromolecular crowding, we address the broader scenario where c_r is not a constant but a function of the reactant concentrations. Thus, we denote c_r as a *stochastic rate function*, while retaining the definition of h_r as above. Knowing that any positive-valued differentiable function can be

approximated locally by a power-law function, we assume the functional form of the stochastic rate function as

$$c_r(\mathbf{x}) = \kappa_r \prod_{s=1}^{N_s} x_s(t)^{\varepsilon_{rs}}. \tag{23}$$

Here, κ_r and ε_{rs} are constants that will be specified in the next section, and $r = 1,\ldots,N_r$. Note that ε_{rs} are now real-valued. Once the stochastic rate function is determined (see below), the propensity function can be calculated as

$$\alpha_r(\mathbf{x}) = c_r(\mathbf{x})h_r(\mathbf{x}) = \frac{\kappa_r}{\prod_{s=1}^{N_s} \underline{\nu}_{rs}!} \prod_{s=1}^{N_s} x_s^{\underline{\nu}_{rs}+\varepsilon_{rs}}. \tag{24}$$

In order to identify the functional expression for a stochastic rate function, and thus the propensity function, we consider the connection between the stochastic and the deterministic equation models. By multiplying CME with **x** and summing over all **x**, we obtain

$$\frac{d}{dt}E\left[\mathbf{X}(t)\right] = \sum_{r=1}^{N_r} \mathbf{v}_r E\left[\alpha_r(\mathbf{X}(t))\right]. \tag{25}$$

Similarly, the expectation for any species $X_s(t)$ is given as

$$\frac{d}{dt}E\left[X_s(t)\right] = \sum_{r=1}^{N_r} \nu_{rs} E\left[\alpha_r(\mathbf{X}(t))\right], \quad s = 1,\ldots,N_s. \tag{26}$$

The details of these derivations are shown in Additional file 1.

We can use these results directly to compute the propensity function for a stochastic GMA model, assuming that its deterministic counterpart is well defined. Specifically, we start with the deterministic GMA equation for X_s,

$$\frac{d}{dt}[X_s(t)] = \sum_{r=1}^{N_r} v_{rs} k_r \prod_{s'=1}^{N_s} [X_{s'}(t)]^{f_{rs'}}, \quad s = 1, \ldots, N_s, \tag{27}$$

where v_{rs}, k_r and $f_{rs'}$ are again the stoichiometric coefficients, rate constants, and kinetic orders, respectively. By substituting $[X_s] = \frac{E[X_s]}{\Phi}$ from Equation (18) into this GMA model, we obtain a "particle-based" equation of the format

$$\frac{d}{dt}\left(\frac{E[X_s]}{\Phi}\right) = \sum_{r=1}^{N_r} v_{rs} k_r \prod_{s'=1}^{N_s} \left(\frac{E[X_{s'}]}{\Phi}\right)^{f_{rs'}}, s = 1, \ldots, N_s .$$

Elementary operations allow us to rewrite this equation as

$$\frac{d}{dt}\left(E[X_s]\right) = \sum_{r=1}^{N_r} v_{rs} k_r \Phi^{1-F_r} \prod_{s'=1}^{N_s} E[X_{s'}]^{f_{rs'}}, \quad s = 1, \ldots, N_s, \tag{28}$$

where $F_r = \sum_{s'=1}^{N_s} f_{rs'}$. In this formulation, the differential operator is justified only when large numbers of molecules are involved. The assumption that the deterministic equations precisely capture the average behaviour of the biochemical reaction system directly equates the stochastic CME (25) to the deterministic equation based model (28)

$$E[\alpha_r(\mathbf{X}(t))] = k_r \Phi^{1-F_r} \prod_{s'=1}^{N_s} E[X_{s'}]^{f_{rs'}}. \tag{29}$$

Now we have two choices for approximating the expectation of the propensity function on left-hand side of equation (29):

1) adopt a zero-covariance assumption as was done in [25], which implies ignoring random fluctuations within every species as well as their correlations. This assumption is only justified for some special cases such as monomolecular and bimolecular reactions under the thermodynamic limit (cf. [4, 6]), but is not necessary valid in generality. Here the thermodynamic limit is defined as a finite

concentration limit which the system reaches when both population and volume approach infinity. Under this assumption, the left hand side of (29) becomes

$$E\left[\alpha_r(\mathbf{x})\right] = E\left[\frac{\kappa_r}{\prod_{s=1}^{N_s} \nu_{rs}!} \prod_{s=1}^{N_s} x_s^{\nu_{rs}+\varepsilon_{rs}}\right] = \frac{\kappa_r}{\prod_{s=1}^{N_s} \nu_{rs}!} \prod_{s=1}^{N_s} E\left[X_s\right]^{\nu_{rs}+\varepsilon_{rs}} \tag{30}$$

for every $r = 1,\ldots,N_r$, and Equation (24) yields

$$\varepsilon_{rs} = f_{rs} - \nu_{rs}$$
$$\kappa_r = k_r \Phi^{1-F_r} \prod_{s=1}^{N_s} \nu_{rs}!$$

$$c_r(\mathbf{x}) = k_r \Phi^{1-F_r} \prod_{s=1}^{N_s} \nu_{rs}! x_s^{\varepsilon_{rs}} \tag{31}$$

and

$$\alpha_{r\_0}(\mathbf{x}) = k_r \Phi^{1-F_r} \prod_{s=1}^{N_s} x_s^{f_{rs}}. \tag{32}$$

Here, the index *r*_0 is used to distinguish this 0-covariance propensity function from a second type of propensity in the next section.

With the zero-covariance assumption, one can substitute (32) back into the equation for the expectation for each species, which yields

$$\frac{d}{dt} E\left[X_s(t)\right] = \sum_{r=1}^{N_r} \nu_{rs} k_r \Phi^{1-F_s} \prod_{s=1}^{N_s} \mu_s^{f_{rs}} \tag{33}$$

for every $s = 1, \ldots, N_s$.. Note that this result is exactly equivalent to the equation-based model (27).

Equation (33) is based on assumption that both the fluctuations within species and their correlations are ignorable, which is not necessarily true in reality. If one uses it in simulations where the assumptions are not satisfied, it is possible that the means for the molecular species are significantly different from the corresponding equation-based model values. This discrepancy arises because the evolution of each species in the stochastic simulation is in truth affected by the covariance which is not necessarily zero, as it was assumed. This phenomenon was observed by Paulsson and collaborators [26] and further discussed in different moment-based approaches [6, 7]. To assess the applicability limit of the propensity defined by (32), we can apply approximation techniques as shown in eqns. (8)-(10) on the functional expression of α_{r_0} and obtain mean and variance as

$$\begin{aligned} \mu_{\alpha_{r\_0}} &= E\left[\alpha_{r\_0}(\mathbf{X}(t))\right] \\ &\approx k_r \Phi^{1-F_r} \prod_{s'=1}^{N_s} E\left[X_{s'}\right]^{f_{rs'}} \exp\left(\sum_{i<j}^{N_s} f_{ri} f_{rj} \operatorname{cov}\left[\log X_i, \log X_j\right]\right) \end{aligned} \tag{34}$$

$$\sigma_{\alpha_{r\_0}}{}^2 \approx \mu_{\alpha_{r\_0}}{}^2 \Omega_r, \tag{35}$$

where

$$\Omega_r = \sum_{s=1}^{N_s} f_{rs} \mu_s^{-2} \sigma_s{}^2 + 2 \sum_{i<j}^{N_s} f_{ri} f_{rj} \operatorname{cov}\left[\log X_i, \log X_j\right], \tag{36}$$

for every $s = 1, \ldots, N_s$. These expressions demonstrate that even with large numbers of molecules the mean of CME does not always converge to the GMA model. Indeed, the convergence is only guaranteed in one of the following special situations: 1) the reaction is of 0^{th} order; 2) the reaction is a real value-order monomolecular reaction, with 1^{st} order reaction as a special case; 3) the covariance contribution in (34) is sufficiently small to be ignored for all participating reactant species of a particular reaction channel. Except for these three special situations, the covariance as shown in (34) significantly affects the mean dynamics. Therefore, stochastic simulations using zero-covariance propensity functions will in general yield means different from what the deterministic GMA model produces. How large these differences are cannot be said in generality. Under the assumption that the GMA model correctly captures the mean dynamics of every species, this conclusion means that α_{r_0} is not necessarily an accurate propensity function for stochastic simulations, and the direct conversion of the equation-based model into a propensity function must be considered with caution.

Moreover, there is no theoretical basis to assume that there are no fluctuations in the molecular species or that these are independent. Therefore, we need to consider the second treatment of the expectation of the propensity function and study the possible effects of a non-zero covariance.

2) We again assume that the GMA model is well defined, which implies that information regarding the species correlations and fluctuations has been captured in the parameters of the GMA model on the left hand size of Equations (7) and (28). To gain information regarding correlations, we use Taylor expansion to approximate the propensity function (see Additional file 1 for details):

$$
\begin{aligned}
E\left[\alpha_r(\mathbf{X}(t))\right] &= E\left[\frac{\kappa_r}{\prod_{s=1}^{N_s}\underline{\nu}_{rs}!}\prod_{s=1}^{N_s}X_s^{\underline{\nu}_{rs}+\varepsilon_{rs}}\right] \\
&\approx \frac{\kappa_r}{\prod_{s=1}^{N_s}\underline{\nu}_{rs}!}\prod_{s=1}^{N_s}E\left[X_s\right]^{\underline{\nu}_{rs}+\varepsilon_{rs}} \\
&\times\exp\left(\sum_{i<j}^{N_s}\left(\underline{\nu}_{ri}+\varepsilon_{ri}\right)\left(\underline{\nu}_{rj}+\varepsilon_{rj}\right)\mathrm{cov}\left[\log X_i,\log X_j\right]\right)
\end{aligned}
\tag{37}
$$

After substitution of (37) in (29), one obtains

$$
\begin{aligned}
\kappa_r &= k_r\Phi^{1-F_r}\prod_{s=1}^{N_s}\underline{\nu}_{rs}!\exp\left(-\sum_{i<j}^{N_s}f_{ri}f_{rj}\,\mathrm{cov}\left[\log X_i,\log X_j\right]\right) \\
\varepsilon_{rs} &= f_{rs}-\underline{\nu}_{rs}.
\end{aligned}
$$

Given the state **x** of the system at time *t***,** the stochastic rate function of reaction R_r is

$$
\begin{aligned}
c_r(\mathbf{x}) &= \kappa_r\prod_{s=1}^{N_s}x_s^{\varepsilon_{rs}} \\
&= k_r\Phi^{1-F_r}\prod_{s=1}^{N_s}\underline{\nu}_{rs}! \\
&\times\exp\left(-\sum_{i<j}^{N_s}f_{ri}f_{rj}\,\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]\right)\prod_{s=1}^{N_s}x_s^{f_{rs}-\underline{\nu}_{rs}}.
\end{aligned}
\tag{38}
$$

Here it is important to understand that although the random variables $\left\{X_s\right\}_{s\in S}$ appear in the expression $c_r(\mathbf{x})$, $c_r(\mathbf{x})$ is not a function of random variables but a deterministic function. The reason is that the $\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]$ in the composition of $c_r(\mathbf{x})$, which as the numerical characteristic of the random variables $\left\{X_s\right\}_{s\in S}$, is deterministic. Therefore, the stochastic rate function $c_r(\mathbf{x})$ is a well-justified deterministic function that is affected by both the state of the system $[x_1,\ldots,x_{N_s}]$ and $\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]$, the numerical characteristic of fluctuations in the random variables $\left\{X_s\right\}_{s\in S}$.

Given the expression $c_r(\mathbf{x})$, the propensity function is

$$\begin{aligned}\alpha_r(\mathbf{x}) &= c_r(\mathbf{x})h_r(\mathbf{x}) \\ &= k_r\Phi^{1-F_r}\prod_{s=1}^{N_s} x_s{}^{f_{rs}} \\ &\times\exp\left(-\sum_{i<j} f_{ri}f_{rj}\,\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]\right).\end{aligned} \tag{39}$$

These results are based on the assumption that there are large numbers of molecules for all reactant species participating in reaction R_r. For simplicity of discussion, we define the *propensity adjustment factor* (paf) of reaction R_r as

$$paf(t) \triangleq \exp\left(-\sum_{i<j} f_{ri}f_{rj}\,\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]\right). \tag{40}$$

paf is a function of time *t* and represents the contribution of the reactants to correlations among species in the calculation of the propensity function for reaction R_r. We denote the propensity function in (39), which accounts for the contribution of the covariance, as α_{r_cov}, in order to distinguish it from the propensity function α_{r_0} (32), which is based on the assumption of zero-covariance, *i.e.*,

$$\alpha_{r\_\mathrm{cov}}(\mathbf{x}) = paf(t)k_r\Phi^{1-F_r}\prod_{s=1}^{N_s} x_s{}^{f_{rs}}. \tag{41}$$

Remembering that $\mathrm{cov}\left[\log X_i(t),\log X_j(t)\right]$, which is a component in both the stochastic rate function $c_r(\mathbf{x})$ and now in the function *paf*(*t*), is a deterministic function rather than a function of random variables, *paf*(*t*) is a deterministic correction to the kinetic constant

k_r in the construction of α_{r_cov} in (41), which corrects the stochastic simulation toward the correct average.

In contrast to the propensity function α_{r_0}, α_{r_cov} leads to accurate stochastic simulations. To illustrate this difference, we analyze $\frac{d}{dt}E\left[X_s(t)\right]$ as follows: We apply the approximation techniques in eqns. (9)-(11) in order to obtain the mean and variance of the propensity function α_{r_cov}:

$$\mu_{\alpha_{r\_\mathrm{cov}}} = E\left[\alpha_{r\_\mathrm{cov}}(\mathbf{X}(t))\right] \approx k_r \Phi^{1-F_r} \prod_{s'=1}^{N_s} E\left[X_{s'}\right]^{f_{rs'}} \tag{42}$$

$$\sigma_{\alpha_{r\_\mathrm{cov}}}{}^2 \approx \mu_{\alpha_{r\_\mathrm{cov}}}{}^2 \Omega_r. \tag{43}$$

Here

$$\Omega_r = \sum_{s=1}^{N_s} f_{rs} \mu_s{}^{-2} \sigma_s{}^2 + 2 \sum_{i<j}^{N_s} f_{ri} f_{rj} \,\mathrm{cov}\left[\log X_i, \log X_j\right]. \tag{44}$$

By substituting (42) back into the derivation of CME (26), one obtains

$$\begin{aligned} &\frac{d}{dt}E\left[X_s(t)\right] \\ &= \sum_{r=1}^{N_r} v_{rs} E\left[\alpha_{r\_\mathrm{cov}}(\mathbf{X}(t))\right] \\ &\approx \sum_{r=1}^{N_r} v_{rs} k_r \Phi^{1-F_r} \prod_{s'=1}^{N_s} \mu_{s'}{}^{f_{rs'}} \end{aligned} \tag{45}$$

for every $s = 1, \ldots, N_s$, which is equivalent in approximation to the GMA model (28). In the other words, the mean of every molecular species obtained by using α_{r_cov} in the

CME derived equation (27) is approximately identical to the corresponding macroscopic variable in the GMA model.

Calculation of $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$

When data in the form of multiple time series for all the reactants are available, it is possible to compute $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$ directly from these data. Once this covariance is known, the function *paf*, α_{r_cov} and the mean dynamics can all be assessed. Alas, the availability of several time series data for all reactants under comparable conditions is rare, so that $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$ must be estimated in a different manner.

If one can validly assume that the covariance based on α_{r_0} does not differ significantly from the covariance based on α_{r_cov}, one may calculate $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$ by one of following methods.

Method 1:

One uses α_{r_0} to generate multiple sets of time series data of all reactants and then computes $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$.

Method 2:

First, $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$ is expressed as a function of mean and covariance in one of the following ways; either as

$$\begin{aligned}\mathrm{cov}\left[\log X_i, \log X_j\right] &\approx \sigma_{ij}/(\mu_i\mu_j) + \frac{1}{2}\log(\mu_i)\left(\sigma_j/\mu_j\right)^2 \\ &+ \frac{1}{2}\log(\mu_j)\left(\sigma_i/\mu_i\right)^2 - \frac{1}{4}\left(\sigma_i/\mu_i\right)^2\left(\sigma_j/\mu_j\right)^2\end{aligned} \tag{46}$$

or as Equation (14):

$$\mathrm{cov}\left[\log X_i, \log X_j\right] = \log\left(1 + \frac{\sigma_{ij}}{\mu_i\mu_j}\right).$$

The first functional expression of $\mathrm{cov}\left[\log X_i(t), \log X_j(t)\right]$ is achieved by Taylor approximation, whereas the second expression is obtained by the additional assumption that the concentrations $\left(X_1, \ldots, X_s\right)$ are log-normally distributed [8, 23]. The consideration of a log-normal distribution is often justified by the fact that many biochemical data have indeed been observed to be log-normally distributed (*e.g.*, [20-22]).

Second, one uses α_{r_0} to approximate the mean and covariance either by direct simulation, as shown in method 1, or by a moment-based approach, which is explained in Additional file 2, and which yields the differential equations

$$\frac{\partial \mu_s}{\partial t} \approx \sum_{r=1}^{N_r} v_{r,s} \left\{ \alpha_{r\_0}(\mathbf{\mu}) + \frac{1}{2} \sum_{m,n=1}^{N_s} \frac{\partial^2 \alpha_{r\_0}(\mathbf{\mu})}{\partial X_m \partial X_n} \sigma_{mn} \right\}$$

$$\frac{\partial \sigma_{ij}}{\partial t} \approx \sum_{r=1}^{N_r} \left\{ v_{r,i} \sum_{s=1}^{N_s} \frac{\partial \alpha_{r\_0}(\mathbf{\mu})}{\partial X_s} \sigma_{js} + v_{r,j} \sum_{s=1}^{N_s} \frac{\partial \alpha_{r\_0}(\mathbf{\mu})}{\partial X_s} \sigma_{is} + v_{r,i} v_{r,j} \left[\alpha_{r\_0}(\mathbf{\mu}) + \frac{1}{2} \sum_{m,n=1}^{N_s} \frac{\partial^2 \alpha_{r\_0}(\mathbf{\mu})}{\partial X_m \partial X_n} \sigma_{mn} \right] \right\}$$

For convenience of computational implementation, the above equations can be written in matrix format

$$\frac{\partial \mu}{\partial t} \approx V^T \left(\alpha + \frac{1}{2} \alpha'' \odot \sigma \right)$$

$$\frac{\partial \sigma}{\partial t} \approx \left(\sigma \left(\alpha' V \right) \right)^T + \sigma \left(\alpha' V \right) + V^T \Lambda V.$$

Here for $r = 1, \ldots, N_r$, and $s, m, n = 1, \ldots, N_s$, $\mu = \left(\mu_1, \ldots, \mu_{N_s} \right)^T$, $\left(V \right)_{rs} = v_{rs}$,

$\alpha = \left(\alpha_1, \ldots, \alpha_{N_r} \right)^T$, $\alpha'' = \left(\alpha_1'', \ldots, \alpha_{N_r}'' \right)^T$, $\left(\alpha_r'' \right)_{mn} = \dfrac{\partial^2 \alpha_r(\mathbf{X})}{\partial X_m \partial X_n}$,

$\alpha_r'' \odot \sigma = \sum_{m,n=1}^{N_s} \dfrac{\partial^2 \alpha_r(\mathbf{X})}{\partial X_m \partial X_n} |_{\mathbf{X}=\mathbf{\mu}} \sigma_{mn}$, $\alpha'' \odot \sigma \triangleq \left(\alpha_1'' \odot \sigma, \ldots, \alpha_{N_r}'' \odot \sigma \right)^T$, $\alpha' = (\alpha_1', \ldots, \alpha_{N_r}')$,

$\alpha_r' = \left(\dfrac{\partial \alpha_r(\mathbf{\mu})}{\partial X_1}, \ldots, \dfrac{\partial \alpha_r(\mathbf{\mu})}{\partial X_{N_s}} \right)^T$, and Λ is a diagonal matrix with

$\left(\Lambda \right)_{rr} = \alpha_r(\mathbf{\mu}) + \dfrac{1}{2} \sum_{m,n=1}^{N_s} \dfrac{\partial^2 \alpha_r(\mathbf{\mu})}{\partial X_m \partial X_n} \sigma_{mn}$.

Statistical criteria for propensity adjustment

Suppose an equation-based model captures the average behavior of a stochastic system and one intends to find the propensity function for a stochastic simulation that will reproduce that means. One can use the 95% confidence interval to evaluate the need for

a propensity adjustment. Specifically, for stable systems that will reach a steady state, we use the reversible reaction model as an example. If the steady state of the ODE x_{st} is within the 95% confidence interval of n runs of stochastic simulations, i.e. $x_{st} \in \left[\mu_{st} - 1.96\frac{\delta_{st}}{\sqrt{n}}, \mu_{st} + 1.96\frac{\delta_{st}}{\sqrt{n}} \right]$, then the rate function in the original ODEs can be used as the propensity without adjustment; otherwise propensity adjustment is needed. Here μ_{st} and δ_{st} can be attained from either a moment-base method or from n independent runs of stochastic simulations using propensity without adjustment. An example discussing a reversible reaction with feedback controls can be found in the results section.

For other systems that do not reach a steady state, but where instead transient characteristics are of the highest interest, one can judge the need of propensity adjustment by whether the pertinent characteristics of the ODEs are within the 95% confidence interval of the corresponding characteristic, which is given by a prediction from the moment-based method or from n runs of stochastic simulations. The Repressilator example in the result section will serve as a demonstration.

Results

Generic special cases

It is generally not valid to translate a rate from a deterministic biochemical model into a propensity function of the corresponding stochastic simulation without adjustment (see Equations. (34)-(36)). However, in some situations, the propensity adjustment (*e.g.*, Equations (40)-(44)) is not needed, and in some other cases it becomes relatively simple.

1) 0^{th}-order reaction kinetics

Consider a very simple equation-based model of the type

$$\frac{d\left[X_s(t)\right]}{dt} = k_r \text{ or } \frac{dE\left[X_s(t)\right]}{dt} = k_r \Phi, \tag{47}$$

for all $s = 1,\ldots,N_s$, $f_{rs} = 0$. According to Equations (40)-(44), one obtains

$$\begin{aligned}
&\Omega_r = 0 \\
&{\sigma_{\alpha_r}}^2 \approx 0 \\
&\mu_{\alpha_r} \approx \exp\left(\log\left(k_r \Phi\right)\right) = k_r \Phi \\
&\text{i.e. } E\left[\alpha_r(\mathbf{X})\right] \approx \alpha_r\left(E\left[\mathbf{X}\right]\right) \\
&\alpha_{r_\text{cov}} \approx k_r \Phi = \alpha_{r_0}.
\end{aligned}$$

Thus, for a 0^{th}-order reaction, its rate equation can be taken directly as the propensity function in stochastic simulations.

2) 1^{st}-order reaction kinetics

Direct application of Equations (40)-(44) yields

$$\begin{aligned}
&\frac{d\left[X_i(t)\right]}{dt} = k_r\left[X_j(t)\right] \\
\text{or } &\frac{dE\left[X_i(t)\right]}{dt} = k_r E\left[X_j(t)\right],
\end{aligned} \tag{48}$$

$f_{rs} = \delta_{sj}$, $i, j = 1,\ldots,N_s$. Therefore, according to Equations (40)-(44)

$$\Omega_r = \left(\sigma_j / \mu_j\right)^2$$
$$\left(\sigma_{\alpha_r} / \mu_{\alpha_r}\right)^2 = \left(\sigma_j / \mu_j\right)^2$$
$$\mu_{\alpha_r} \approx \exp\left(\log\left(k_r \mu_j\right)\right) = k_r \mu_j$$
i.e. $E\left[\alpha_r(\mathbf{X})\right] \approx \alpha_r\left(E\left[\mathbf{X}\right]\right)$
$$\alpha_{r\_\mathrm{cov}}(\mathbf{X}) \approx k_r X_j = \alpha_{r\_0}(\mathbf{X}).$$

Thus, for 1st-order reactions, the rate equation can again be taken directly as the propensity function in stochastic simulations.

3) Real-valued order monomolecular reaction kinetics

Consider a reaction with kinetics of the type

$$\frac{d\left[X_i(t)\right]}{dt} = k_r\left[X_j(t)\right]^{f_{rj}}$$
$$\text{or } \frac{dE\left[X_i(t)\right]}{dt} = k_r \Phi^{1-f_{rj}} E\left[X_j(t)\right]^{f_{rj}}, \tag{49}$$

$f_{rj} \neq 0$, $f_{rs} = 0$, for any $s \neq j, s = 1, \ldots, N_s$. Equations (40)-(44) lead to

$$\Omega_r = \left(\sigma_j / \mu_j\right)^2$$
$$\left(\sigma_{\alpha_r} / \mu_{\alpha_r}\right)^2 = \left(\sigma_j / \mu_j\right)^2$$
$$\mu_{\alpha_r} \approx k_r \Phi^{1-f_{rj}} \mu_j^{\,f_{rj}} \text{ i.e. } E\left[\alpha_r(\mathbf{X})\right] \approx \alpha_r\left(E\left[\mathbf{X}\right]\right)$$
$$\alpha_{r\_\mathrm{cov}}\left(\mathbf{X}\right) \approx k_r \Phi^{1-f_{rj}} X_j^{\,f_{rj}} = \alpha_{r\_0}\left(\mathbf{X}\right).$$

Thus, for reaction kinetics involving a single variable and a real-valued order, the rate equation can again be taken as the propensity function in stochastic simulations.

4) 2nd-order reaction kinetics

This type of reaction can be expressed as

$$\begin{aligned} &\frac{d}{dt}\left[X_s(t)\right] = k_r\left[X_i(t)\right]\left[X_j(t)\right] \\ \text{or } &\frac{dE\left[X_s(t)\right]}{dt} = k_r\Phi^{-1}E\left[X_i(t)\right]E\left[X_j(t)\right], \end{aligned} \tag{50}$$

$i, j \in \{1,\ldots,N_s\}, i \neq j$, $f_{ri} = f_{rj} = 1$, and $f_{rs} = 0$, for all $s \neq i, j$. Therefore, according to Equations (40)-(44)

$$\begin{aligned} \Omega_r &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_j/\mu_j\right)^2 + 2\,\mathrm{cov}\left[\log X_i, \log X_j\right] \\ &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_s/\mu_s\right)^2 + 2\left\{\mathrm{cov}\left[X_i/\mu_i, X_j/\mu_j\right]\right. \\ &\left. + \frac{1}{2}\log(\mu_i)\left(\sigma_j/\mu_j\right)^2 + \frac{1}{2}\log(\mu_j)\left(\sigma_i/\mu_i\right)^2 - \frac{1}{4}\left(\sigma_i/\mu_i\right)^2\left(\sigma_j/\mu_j\right)^2\right\} \end{aligned}$$

$$\begin{aligned} \left(\sigma_{\alpha_r}/\mu_{\alpha_r}\right)^2 &= \Omega_r \\ \partial &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_j/\mu_j\right)^2 + 2\,\mathrm{cov}\left[\log X_i, \log X_j\right] \\ \mu_{\alpha_r} &\approx k_r(N_A V)^{-1}\mu_i\mu_j \\ \alpha_{r\_\mathrm{cov}}(\mathbf{X}) &= k_r\Phi^{-1}X_iX_j\exp\left(-\mathrm{cov}\left[\log X_i, \log X_j\right]\right) \neq \alpha_{r\_0}(\mathbf{X}). \end{aligned}$$

Thus, the proper propensity function for 2nd-order reactions is different from the rate equation. The difference can be ignored only if the contribution from the covariance is insignificant. In general, the rate equation yields only an approximate propensity function for stochastic simulations, and the approximation quality must be assessed on a case-by-case basis.

5) Bimolecular reaction with real-valued order kinetics

This type of reaction can be formulated as

$$\begin{aligned} &\frac{d\left[X_s(t)\right]}{dt} = k_r\left[X_i(t)\right]^{f_i}\left[X_j(t)\right]^{f_j} \\ \text{or } &\frac{dE\left[X_s(t)\right]}{dt} = k_r\Phi^{1-f_i-f_j}E\left[X_i(t)\right]^{f_i}E\left[X_j(t)\right]^{f_j}, \end{aligned} \tag{51}$$

$i, j \in \{1,\ldots,N_s\}, i \neq j$, f_{ri}, $f_{rj} \neq 0$, and $f_{rs} = 0$, for all $s \neq i, j$. According to Equations (40)-(44) we obtain

$$\begin{aligned}\Omega_r &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_j/\mu_j\right)^2 + 2 f_i f_j \operatorname{cov}\left[\log X_i, \log X_j\right] \\ &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_j/\mu_j\right)^2 + 2 f_i f_j \left\{\operatorname{cov}\left[X_i/\mu_i, X_j/\mu_j\right]\right. \\ &\left. + \frac{1}{2}\log(\mu_i)\left(\sigma_j/\mu_j\right)^2 + \frac{1}{2}\log(\mu_j)\left(\sigma_i/\mu_i\right)^2 - \frac{1}{4}\left(\sigma_i/\mu_i\right)^2\left(\sigma_j/\mu_j\right)^2\right\}\end{aligned}$$

$$\begin{aligned}\left(\sigma_{\alpha_r}/\mu_{\alpha_r}\right)^2 &= \Omega_r \\ &= \left(\sigma_i/\mu_i\right)^2 + \left(\sigma_j/\mu_j\right)^2 + 2 f_i f_j \operatorname{cov}\left[\log X_i, \log X_j\right] \\ \mu_{\alpha_r} &\approx k_r \Phi^{f_i + f_j - 1} \mu_i^{f_i} \mu_j^{f_j} \\ \alpha_{r\_\mathrm{cov}}(\mathbf{X}) &= k_r \Phi^{f_i + f_j - 1} X_i^{f_i} X_j^{f_j} \exp\left(-f_i f_j \operatorname{cov}\left[\log X_i, \log X_j\right]\right) \\ &\neq \alpha_{r\_0}(\mathbf{X}).\end{aligned}$$

For bimolecular reactions of complex order, the propensity function is different from the rate equation. The difference can be ignored only if the contribution from the covariance is insignificant.

Power-law representation of a reversible reaction with feedback controls

We consider a reversible reaction with feedback controls (see Figure 1) whose average behaviour is accurately described by the following GMA model

$$\begin{aligned}\frac{dx_1}{dt} &= \frac{dx_2}{dt} = -\frac{dx_3}{dt} \\ &= -k_f \Phi^{1-f_1-f_2-f_3} x_1^{f_1} x_2^{f_2} x_3^{f_3} + k_b \Phi^{1-g_1-g_3} x_1^{g_1} x_3^{g_3}.\end{aligned} \tag{52}$$

Here S_3 feeds back to inhibit the forward reaction and S_1 feeds back on the reverse reaction and accelerates it. The task is to develop a stochastic model whose performance converges to that of the deterministic GMA model. We can see from equations (52) that three variables x_1, x_2 and x_3 contribute to the forward flux $k_f \Phi^{1-f_1-f_2-f_3} x_1^{f_1} x_2^{f_2} x_3^{f_3}$ and two variables x_1 and x_3 contribute to the backward flux $k_b \Phi^{1-g_1-g_3} x_1^{g_1} x_3^{g_3}$. Because several

variables are involved, their covariance has the potential of affecting the forward and the backward propensity functions in a stochastic simulation. To obtain the covariance information, we formulate the moment equations (53) from the ODE model (52).

To simplify the calculation, as explained in detail in Additional file 2, we set the third central moment to zero and obtain a closed-form set of ODEs. Expressed differently, the rate of change in mean and covariance depends only on the functions of mean and covariance themselves, but not on higher-order moments. Thus,

$$\begin{aligned} \frac{\partial \mu}{\partial t} &\approx V^T \left(\alpha + \frac{1}{2} \alpha'' \odot \sigma \right) \\ \frac{\partial \sigma}{\partial t} &\approx \left(\sigma \left(\alpha' V \right) \right)^T + \sigma \left(\alpha' V \right) + V^T \Lambda V. \end{aligned} \tag{53}$$

Here $\mu = \left(\mu_1, \mu_2, \mu_3 \right)^T$, $V = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$, $\alpha = \left(\alpha_1, \alpha_2 \right)^T = \begin{bmatrix} k_f \Phi^{1-f_1-f_2-f_3} x_1^{f_1} x_2^{f_2} x_3^{f_3} \\ k_b \Phi^{1-g_1-g_3} x_1^{g_1} x_3^{g_3} \end{bmatrix}$.

Moreover, for $r = 1,2$ and $m,n = 1,2,3$, $\left(\alpha_r'' \right)_{mn} = \dfrac{\partial^2 \alpha_r(\mathbf{X})}{\partial x_m \partial x_n}$, $\alpha'' = \left(\alpha_1'', \alpha_2'' \right)^T$,

$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$, $\alpha_r'' \odot \sigma \triangleq \sum_{m,n=1}^{3} \dfrac{\partial^2 \alpha_r(\mathbf{X})}{\partial x_m \partial x_n} |_{\mathbf{X}=\boldsymbol{\mu}} \, \sigma_{mn}$, $\alpha'' \odot \sigma \triangleq \left(\alpha_1'' \odot \sigma, \alpha_2'' \odot \sigma \right)^T$,

$\alpha' = (\alpha_1', \alpha_2')$, $\alpha_r' = \left(\dfrac{\partial \alpha_r(\boldsymbol{\mu})}{\partial x_1}, \dfrac{\partial \alpha_r(\boldsymbol{\mu})}{\partial x_2}, \dfrac{\partial \alpha_r(\boldsymbol{\mu})}{\partial x_3} \right)^T$, and

$$\Lambda = \begin{bmatrix} \alpha_1(\boldsymbol{\mu}) + \dfrac{1}{2} \displaystyle\sum_{m,n=1}^{3} \dfrac{\partial^2 \alpha_1(\boldsymbol{\mu})}{\partial x_m \partial x_n} \sigma_{mn} & 0 \\ 0 & \alpha_2(\boldsymbol{\mu}) + \dfrac{1}{2} \displaystyle\sum_{m,n=1}^{3} \dfrac{\partial^2 \alpha_2(\boldsymbol{\mu})}{\partial x_m \partial x_n} \sigma_{mn} \end{bmatrix}.$$

Two initial conditions are chosen for representative simulations; they differ by a factor of 20 in species populations and reaction volume between the upper and lower panels of Figure 2. The purpose is to observe the thermodynamic limit of the systems: both scenarios have the same initial concentrations, but the system in the lower panel case has a larger species populations and reaction volume and can thus be regarded as the

thermodynamic limit sample of system in the upper panel. As demonstrated by the figures in the first column, the moment approach predicts that for both population sizes the average trajectories of the stochastic model (without propensity adjustment) dynamics is lower than that of the equation-based model: the differences are about 10% of the steady-state value of the equation-based model in the upper figure and 1% in the lower figure; for 100 runs of the stochastic simulation, the steady-state value of the equation-based model lies outside the 95% confidence interval in the upper figure, while it is inside the interval in the lower figure. Therefore, we can expect that the propensity adjustment will significantly contribute to the stochastic simulation for the upper case while not for the lower case. This expectation is confirmed by the simulation results in the third and fourth columns. With the common assumption that the deterministic equations precisely capture the system's average behaviour, the case in the upper panel represents the situation where propensity adjustment is needed, while the lower panel represents the situation that a propensity without adjustment is sufficient when the system approaches its thermodynamics limit. This example furthermore demonstrates that either the moment approach or the stochastic simulations without propensity adjustment can be used to estimate whether there is a need to construct a propensity adjustment function for stochastic simulations.

Repressilator

Interestingly, a propensity function may even be obtained through power-law approximation of some function that describes complex transient behaviours of a reaction network. As an example, consider the so-called *Repressilator* [27], which is a three-component genetic circuit where each component represses its downstream neighbour. More specifically (as shown in Figure 3), gene G_1 codes for protein x_1, whose dimer y_1

subsequently represses the transcription of the gene G_2. Similarly, y_2, the dimer of gene G_2's protein product x_2, represses the transcription of gene G_3, and y_3, the dimer of gene G_3's protein product x_3, represses the transcription of gene G_1. The corresponding differential equation model following mass action kinetics is given by [28]

$$\begin{aligned} x_i' &= -2\kappa_+ x_i^2 + 2\kappa_- y_i + \sigma m_i - \gamma_p x_i \\ y_i' &= \kappa_+ x_i^2 - \kappa_- y_i - k_+ y_i d_{0,j} + k_- d_{r,j} \\ d_{0,i}' &= -k_+ y_k d_{0,i} + k_- d_{r,i} \\ d_{r,i}' &= k_+ y_k d_{0,i} - k_- d_{r,i} \\ m_i' &= d_{0,i} - \gamma_m m_i, \end{aligned} \tag{54}$$

where $i = 1,2,3;\ j = 2,3,1; k = 3,1,2;$ the rate constants are explained in the diagram below

$$\begin{aligned} x_i + x_i &\underset{\kappa_-}{\overset{\kappa_+}{\rightleftharpoons}} y_i \\ d_{0,i} + y_k &\underset{k_-}{\overset{k_+}{\rightleftharpoons}} d_{r,i} \\ d_{0,i} &\xrightarrow{\alpha} d_{0,i} + m_i \\ m_i &\xrightarrow{\sigma} m_i + x_i \\ x_i &\xrightarrow{\gamma_p} \phi \\ m_i &\xrightarrow{\gamma_m} \phi \end{aligned}$$

Assuming that the reversible dimerization and the dissociation/association of a protein dimer from/to the promoter are much faster than other processes, the full systems can be reduced to

$$\begin{aligned} x_i' &= \sigma p(x_i)^{-1} m_i - \gamma_p p(x_i)^{-1} x_i \\ m_i' &= \frac{\alpha d}{1 + c_d c_p x_k^2} - \gamma_m m_i \end{aligned} \tag{55}$$

[28]. Here $\Phi = 1$, $p(x_i) = 1 + 4c_p x_i + \frac{4c_d c_p d x_i}{(1 + c_d c_p x_i^2)^2}$, $c_p = \kappa_+ / \kappa_-$, $c_d = k_+ / k_-$ and $d = d_{0,i} + d_{r,i}$

for *i*=1,2,3. It has been shown that the simplified ODEs rather accurately approximate the

transient dynamics of the full system by retaining the original oscillation period and amplitude.

In [28], the system (55) is further rescaled by setting $\tilde{t} = \gamma_m t$, $\tilde{x}_i = \sqrt{c_d c_p}\, x_i$ and $\tilde{m}_i = \left(\sigma\sqrt{c_d c_p}\, m_i\right) \Big/ \left(\gamma_m \beta\right)$, which yields

$$\begin{aligned} \frac{d\tilde{x}_i}{d\tilde{t}} &= \beta\, p(\tilde{x}_i)^{-1}\tilde{m}_i - \beta\, p(\tilde{x}_i)^{-1}\tilde{x}_i \\ \frac{d\tilde{m}_i}{d\tilde{t}} &= \frac{\kappa d'}{1+\tilde{x}_k^{\,2}} - \tilde{m}_i. \end{aligned} \tag{56}$$

Intriguingly, one makes the following observation. The scaled ODE system (56) is consistent with the original system (55) in oscillation amplitude and period. However, its corresponding stochastic model produces results that deviate substantially from the average responses. To see the effects of the transition from a deterministic to a stochastic model, we apply SSA to the scaled system (56). The main result is that the oscillation periods of both x_i and m_i are reduced to half (Figure 4). The reason is that, in the stochastic simulation, the oscillation period is very sensitive to the ratio of x_i and m_i, which has been altered by the scaling operation. Therefore, in general one needs to pay attention to how scaling may affect the stochastic performance when the model is generated through the conversion of an ODE model.

We can see from equations (55) that two variables x_i and m_i contribute to the production of x_i; hence, their covariance could affect the propensity function of x_i in the production reaction of a stochastic simulation. Similar to the example of a reversible reaction (Equation 52), it is therefore necessary to evaluate covariance effects and to judge whether the propensity function needs adjusting. Thus, we need to compare the difference

between the dynamics of the phenomenological model (55) and the dynamics under the influence of covariance, which can be produced by either stochastic simulation or the moment approach.

The influence of the covariance on the dynamics of the stochastic simulation is relatively easy to assess: we simply use the terms on the right-hand side of the differential equations (54) as the propensity functions in SSA and obtain simulation results shown in the 2nd and the 4th panels of Figure 5. Obtaining the covariance-influenced dynamics with the moment-based approach is more complicated, and we need to discuss some implementation issues.

First, the moment-based approach requires information regarding the first and the second derivatives of $p(x_i)^{-1}$, which have rather complicated functional forms. To simplify the calculation, we replace the function $p(x_i)^{-1}$ with an approximating power-law function. Specifically, suppose the original parameter values are $\kappa_+ = k_+ = 5$, $\kappa_- = k_- = 100$ and $d = 20$. Plotting the data $\left(x_i, p(x_i)^{-1}\right)$ in log-log space (Figure 5) indicates that the original function is represented well by a straight line:

$$\log y_i = \log 3.5188 - 0.9384 \log x_i \quad .$$

for $x_i \in [30, 300]$. In Cartesian space, this line corresponds to the power-law function

$$y_i = 3.5188 x_i^{-0.9384},$$

which models the original function very well (see Figure 5). For $x_i \in [1, 30]$, this power-law function does not fit the original function precisely; the effect of this imprecision can be evaluated later at after we use this power-law function in the moment-based method.

Moreover, using the truncated moment equations to estimate the mean and variance involves multiple approximations: First, the function $p(x_i)^{-1}$ on the right-hand side of (55) is replaced by a power-law function (see Figure 5). Second, the result is approximated by Taylor expansion to the second order. Third, similar to the example of a reversible reaction, the central moment of the third degree is assumed to be zero, which leads to a closed-form ODE for the first two moments.

Solving the technical issues as described, one obtains the corresponding moment-based model of (55) (not shown) with results shown in Figure 6. Suppose one is particularly interested in the period and amplitude of the oscillation within a time interval between 0 and 400 seconds. As shown in Figure 6, the GMA approximation (black dashed line) fits the original ODEs (55) (bold black solid line) very well at the beginning, but as time goes on, the approximation error accumulates. As seen in the time interval [350, 400], the GMA approximation deviates from the original ODEs significantly. However, this does not mean that the GMA approximation cannot be used as a propensity function for stochastic simulations; the moment-based method with the GMA approximation shows that, when the GMA approximation is used as propensity function (without adjustment) for stochastic simulations, the resulting mean (red solid line) consistently fits the trajectory of the original ODEs (bold solid black line) very well up to about t=400 seconds. The oscillation period and amplitude in the stochastic simulation based on the GMA approximation (without adjustment) are almost identical to those of the original ODEs. Therefore, a propensity adjustment for the GMA approximation is not needed, and the GMA approximation can be used as a propensity function for stochastic simulations. In other words, a stochastic model for the Repressilator system can be generated by using the scheme in (32) without propensity adjustment. Moreover, the imprecision caused by

the power-law approximation can be tolerated when its corresponding moment-based mean matches the original ODEs sufficiently well with respect to the features of highest interest.

Enzymatic reaction using a quasi-steady state assumption (QSSA)

We consider an enzymatic reaction following the Michaelis-Menten mechanism:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} P + E.$$

Here enzyme E reacts with substrate S through a reversible reaction to form complex ES, which can proceed to yield product P and to release the enzyme E. By assuming the law of mass action for the reaction kinetics we obtain a set of differential equations for the system dynamics:

$$\begin{aligned} \frac{d[S]}{dt} &= k_{-1}[ES] - k_1[S]\left([E]_0 - [ES]\right) \\ \frac{d[ES]}{dt} &= k_1[S]\left([E]_0 - [ES]\right) - \left(k_1 + k_2\right)[ES] \\ \frac{d[P]}{dt} &= k_2[ES], \end{aligned} \tag{57}$$

where the total amount of enzyme in the form of free enzyme and complex $[E]_0 \triangleq [E] + [ES]$ is assumed to be constant. In addiction, by making the so-called *quasi steady state assumption* (QSSA) [29, 30], assuming that the complex ES is essentially in steady state, we can assert $\frac{d[ES]}{dt} \approx 0$. As it has been discussed many times in the literature, QSSA reduces the system and leads to the approximate form

$$\begin{aligned} \frac{d[S]}{dt} &= -\frac{V_{\max}[S]}{K_m + [S]} \\ \frac{d[P]}{dt} &= \frac{V_{\max}[S]}{K_m + [S]}, \end{aligned} \tag{58}$$

which is known as *Michaelis-Menten kinetics* [30]. The characterizing parameters are $V_{\max} = k_2[E]_0$ and $K_m = (k_{-1} + k_2)/k_1$.

Applying QSSA, Rao and Arkin [2] were able to reduce the CME of S and ES to a CME only containing S. For the reduced CME, the propensity function for the overall reaction $S \to P$ is

$$\alpha(s) = \frac{V_{\max} s}{K_m + s}, \tag{59}$$

where the volume was scaled so that $\Phi = 1$ and the lower-case letter *s* denotes the molecule count of species S. Instead of reviewing the relatively complicated manipulations with CME, we show in the following that the techniques described above lead directly from the equation-based model to the propensity function for the reduced system.

First, we recast the equation-based model into the GMA format [31], by introducing an auxiliary variable $[T] \triangleq K_m + [S]$. The result,

$$\begin{aligned} \frac{d[S]}{dt} &= -V_{\max}[S][T]^{-1} \\ \frac{d[T]}{dt} &= \frac{d[S]}{dt} = -V_{\max}[S][T]^{-1} \end{aligned} \tag{60}$$

is exactly equivalent to the reduced system in (58) with the initial condition $[S]_0$ and $[T]_0 = K_m + [S]_0$. The corresponding stochastic model has only one reaction channel and the propensity function is

$$\alpha(s,t) = V_{\max} s t^{-1}. \tag{61}$$

The propensity adjustment factor can be set to 1 because T is a function of s and its covariance with s is therefore 1. By applying $t = K_m + s$, the propensity function can be simplified as

$$\alpha(s,t) = V_{\max} s t^{-1} = V_{\max} s (K_m + s)^{-1} = \alpha(s). \qquad (62)$$

Thus, we arrive at the propensity function for the reduced system, which is identical to the result of Rao and Arkin obtained through manipulations of CME.

In the above derivation, we used the simplest type of recasting, where a new, auxiliary variable simply consists of an old variable plus a constant. This reformulation of the Michaelis-Menten process as a pair of GMA equations is a special case of a much more general recasting technique that permits the equivalent conversion of any system of ordinary differential equations into a power-law format [31]. However, this equivalence transformation imposes constraints on the variables of the GMA equations, and it is at this point unclear whether there are mathematical warranties ensuring that the proposed transition from differential to stochastic equations in general preserves these constraints in all cases. This question will require further investigation.

Stochastic Focusing

Stochastic focusing [26] describes the phenomenon that the fluctuations of a chemical species can drive the system to reach a different steady state than what a deterministic ODE model predicts. To demonstrate the utility of propensity adjustment, we derive a stochastic model which produces consistent results with those of the deterministic model.

Following [32], we consider the following reactions system

$$\begin{aligned} &\phi \xrightarrow{k_1} I \xrightarrow{k_2} P \xrightarrow{k_3} \phi \\ &I + S \xrightarrow{k_4} S \\ &S \underset{k_5}{\overset{k_6}{\rightleftharpoons}} \phi \end{aligned} \tag{63}$$

This system can be interpreted as follows: the intermediate species I is produced at constant rate k_1 from some source ϕ and degrades with rate k_4 through the catalysis with signalling molecule S; the end product P is converted from species I at rate k_2 and degrades at rate k_3; the signalling molecule S is produced and degrades at rates k_5 and k_6, respectively. Moreover, the value of k_5 is reduced to half at a certain time point to achieve a significant divergence effect. In order to capture the average dynamics of the system accurately, we use a power-law model in GMA format instead of the mass action rate law in [32].

$$\begin{aligned} \frac{di}{dt} &= k_1 - k_2 i - k_4 i^{f_I} s^{f_S} \\ \frac{dp}{dt} &= k_2 i - k_3 p \\ \frac{ds}{dt} &= k_5 - k_6 s \end{aligned} \tag{64}$$

The system size is set to 1. We can see from equations (64) that two variables i and s contribute to the degradation of I and that their covariance could therefore affect the propensity function of I in the degradation reaction of a stochastic simulation. To calculate the propensity adjustment function $paf_4(t) = \exp\left(-f_I f_S \operatorname{cov}\left[\log I(t), \log S(t)\right]\right)$ for reaction $R_4 : I + S \xrightarrow{k_4} S$, we formulate equations (*cf.* (60)) for the moments as

$$\begin{aligned} \frac{\partial \mu}{\partial t} &\approx V^T \left(\alpha + \frac{1}{2} \alpha'' \odot \sigma \right) \\ \frac{\partial \sigma}{\partial t} &\approx \left(\sigma\left(\alpha' V\right) \right)^T + \sigma\left(\alpha' V\right) + V^T \Lambda V. \end{aligned} \tag{65}$$

Here $\mu = \left(\mu_I, \mu_P, \mu_S\right)^T$, $V = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix}$, $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 i \\ k_3 p \\ k_4 i^{f_I} s^{f_S} \\ k_5 \\ k_6 s \end{bmatrix}$.

Moreover, for $r = 1,\ldots,6$ and $m,n = i,p,s$, $\left(\alpha_r''\right)_{mn} = \dfrac{\partial^2 \alpha_r}{\partial m \partial n}$, $\alpha'' = \left(\alpha_1'', \ldots, \alpha_6''\right)^T$,

$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$, $\alpha_r'' \odot \sigma = \sum_{m,n=i,p,s} \dfrac{\partial^2 \alpha_r(\boldsymbol{\mu})}{\partial m \partial n} \sigma_{mn}$, $\alpha'' \odot \sigma \triangleq \left(\alpha_1'' \odot \sigma, \ldots, \alpha_6'' \odot \sigma\right)^T$,

$\alpha' = (\alpha_1', \ldots, \alpha_6')$, $\alpha_r' = \left(\dfrac{\partial \alpha_r}{\partial i}, \dfrac{\partial \alpha_r}{\partial p}, \dfrac{\partial \alpha_r}{\partial s}\right)^T$, and the diagonal matrix Λ is defined by $\left(\Lambda\right)_{rr} = \alpha_r + \dfrac{1}{2} \sum_{m,n=i,p,s} \dfrac{\partial^2 \alpha_r}{\partial m \partial n} \sigma_{mn}$.

The stochastic focusing model without propensity adjustment yields results quite different from those of the deterministic model, as is illustrated in Figure 7. In this figure, the blue lines in the 1st panel are predicted from the moment equations (65) and the blue error bars for μ_P in the 2nd panel are obtained from ten independent stochastic simulations. Both diverge systematically from the black line predicted by ODE model (64). By contrast, the stochastic model with propensity adjustment produces results consistent with the deterministic model, as shown by the 4th panel.

Discussion

It is often implicitly assumed that the rate of a dynamic process can be directly taken as the propensity for a corresponding stochastic process. We have shown here that this is sometimes, but not always, true. Our results fall into three categories. The first develops conditions for a valid conversion of a rate to a propensity, the second presents a general

conversion procedure, and the third discusses computational issues of propensity adjustment.

Conditions for the direct use of a rate constant (function) as propensity function

We have shown that the direct use of a rate constant or a rate function *f* as the propensity function in a stochastic simulation algorithm requires that at least one of the following assumptions be true:

1) *f* is a linear function; this assumption has been validated in the Results sections addressing 0^{th}-order and 1^{st}-order reaction kinetics.

2) the reaction is monomolecular; this assumptions was evaluated in the Results section describing real-valued order monomolecular reaction kinetics.

3) all X_i in the system are noise-free variables, *i.e.*, without (or with ignorable) fluctuations; this assumption implies that the covariance of any two participating reactants is zero (or close to zero). This assumption is assessed in equations (29 - 36).

Each of these three conditions is a sufficient condition for the direct use of a rate function *f* as the propensity function. Moreover, these statements are valid for functions of a general format, not just for GMA. This is so because the functional formats in cases 1 and 2 above are special cases of the GMA format. For the third case, a formal proof is only given for functions in GMA format, because this structured format allows us to give an explicit estimation on how the covariance can affect the average behavior of a stochastic simulation through equation (34). For functions not in GMA format, the conclusion is still holds, although an analogous explicit estimation is lacking. The argument is as

follows. The bimolecular reaction $E\left[\alpha_r(\mathbf{X}(t))\right]$ contains at least one quadratic moment of the form $E\left[X_i(t)X_j(t)\right]$ (*cf.* [4] and page 38). Therefore, by definition of the covariance, $E\left[X_i(t)X_j(t)\right] = E\left[X_i(t)\right]E\left[X_i(t)\right] + \mathrm{cov}\left(X_i(t), X_j(t)\right)$, we obtain

$$E\left[X_i(t)X_j(t)\right] = E\left[X_i(t)\right]E\left[X_i(t)\right] \Leftrightarrow \ \mathrm{cov}\left(X_i(t), X_j(t)\right) = 0\,.$$

This result implies the following: If the covariance between every pair of random variables is zero (or ignorable), we have $E\left[X_i(t)X_j(t)\right] = E\left[X_i(t)\right]E\left[X_i(t)\right]$ and therefore $E\left[\alpha_r(\mathbf{X}(t))\right] = \alpha_r(E\left[\mathbf{X}(t)\right])$. Expressed in words, the expectation of the propensity function on left-hand side of equation (29) equals its rate function, and the rate function can be directly used as propensity function in stochastic simulations.

If at least one of the three assumptions is satisfied, the stochastic simulation algorithm (SSA) is applicable without changes.

A general procedure for converting an equation-based model into a stochastic analogue

In the past, efforts have been made to manipulate the chemical master equation (CME) in order to achieve a proper propensity function for a reduced system (*e.g.*, see [2]). However, manipulations of CME are usually complicated, and successes have been modest and rare. Here we propose an alternative strategy for converting a reduced dynamical model into a stochastic analogue. To achieve this conversion, we addressed two fundamental issues: First, under what conditions can a deterministic, equation-based model be validly used in stochastic simulations? And second, what is a proper strategy to implement such a conversion?

To address the first question, we showed that the following steps are necessary:

(1) A concentration-based model needs to be converted into a particle-based model by accounting for the size of the system; if the concentration-based model is scaled (as was illustrated with the repressilator example), it may first have to be un-scaled in order to render the conversion valid;

(2) The difference between the mean of a stochastic model without propensity adjustment and the corresponding quantities of the equation-based model should be evaluated. The mean of the stochastic model is obtained either through stochastic simulations or through a moment-based approach. If the difference is significant, then an adjustment of the propensity function for a non-elementary reaction is necessary.

To answer the second question, we need to execute the following steps

(3) Compute a propensity adjustment function, either through simulated or experimental data or through a moment-based approach, in order to achieve the corrected propensity function (41);

(4) Apply SSA or one of its variants using a propensity function with adjustment to obtain valid simulation trajectories.

Computational issues of propensity adjustments

When the propensity needs adjusting, an accurate propensity adjustment function (*paf*) is essential for obtaining the proper correction of the propensity. It is usually impossible to compute *paf* exactly, which necessitates a suitable approximation. The approximation error in *paf* originates from the following sources:

1) The expression of *paf* in Equation (40) is a function of the mean, variance, and covariance, which are computed with a 2^{nd}-order Taylor expansion in log space.

2) The moment-based approach, from which the functions of mean, variance and covariance are usually derived, is an approximation method that yields a closed ODE system for the moments. In the method used here, the propensity function is approximated by a 2^{nd}-order Taylor expansion, and the moments up to a certain degree (2 in our treatment) are retained, while all higher moments are assumed to be zero. One might expect that a higher-order Taylor expansion would improve the accuracy of *paf*, but it would come with a much higher computational cost. The error control of *paf* and the relative computational issues should be addressed in future studies.

Since computation cost is a major concern with the stochastic simulation of large biochemical reaction networks, another issue has yet to be addressed. Namely, how does the propensity function of a reduced system affect the accuracy and efficiency of various leaping methods that have been proposed to speed up SSA? Moreover, the question of molecular population sizes requires further analysis. Our derivation assumed large reactant populations, but simulations of a reversible pathway indicated that the method works rather well even for small populations. A more careful investigation of this issue of population size in different scenarios is still needed and should be the subject of further research.

Conclusions

Gillespie's stochastic simulation algorithm (SSA), as well as later variants, permits three kinds of elementary reactions to be modelled: 0^{th}, 1^{st} and 2^{nd} order reactions that are assumed to follow the law of mass action. All other types of reactions, containing non-integer kinetic orders and/or following other types of kinetic law, are assumed to be

convertible to one of these three kinds, so that SSA can validly be applied. However, the conversion to elementary reactions is often difficult, infeasible, or simply impossible. First, the kinetic parameters of the underlying elementary reactions are in many cases unknown for a complex-order reaction. Second, even when all elementary kinetic parameters are available, the multitude of reaction channels and participating species creates a combinatorial complexity that renders SSA simulations computationally impractical. Within a deterministic framework, model reduction is a possible and often-used strategy to address such challenges. For example, a reduced mechanistic model, such as the Michaelis-Menten rate law, is often proposed to fit the experimental data, at the cost of sacrificing the original mechanistic interpretation. The reduction in these cases simplifies the original formulation by approximating, merging, or omitting intermediate reaction steps and reactants.

In this article, we propose a rather general strategy for converting a deterministic process model into a corresponding stochastic model and characterize the mathematical connections between the two. The deterministic framework is assumed to be a generalized mass action system and the stochastic analogue is in the format of the chemical master equation. The analysis identifies situations: where a direct conversion is valid; where internal noise affecting the system needs to be taken into account; and where the propensity function must be mathematically adjusted. The conversion from deterministic to stochastic models is illustrated with several representative examples, including reversible reactions with feedback controls, Michaelis-Menten enzyme kinetics, a genetic regulatory motif, and stochastic focusing. The construction of a stochastic model for a biochemical network requires the utilization of information associated with an equation-

based model. The conversion strategy proposed here guides a model design process that ensures a valid transition between deterministic and stochastic models.

Authors' contributions

JW developed the mathematical derivations, designed and performed the simulation, and drafted the manuscript. BV contributed to the statistical reasoning and revised the manuscript. EV supervised the research and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Dr. Yi Jiang for useful comments and for providing seminal references. The authors also appreciate Dr. Mukhtar Ullah's and Dr. Olaf Wolkenhauer's insightful comments and conceptual clarifications. This work was supported in part by a Molecular and Cellular Biosciences Grant (MCB-0946595; E.O. Voit, PI) from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring institutions.

References

1. Gillespie, D., *Exact Stochastic Simulation of Coupled Chemical Reactions.* J Phys Chem, 1977. 81(25): p. 2340 - 2361.

2. Rao, C.V. and A.P. Arkin, *Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm.* The Journal of chemical physics, 2003. 118(11): p. 4999-5010.

3. Cao, Y., D.T. Gillespie, and L.R. Petzold, *Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems.* J. Comput. Phys., 2005. 206: p. 395.

4. Gillespie, D.T., *Stochastic simulation of chemical kinetics.* Annual Review of Physical Chemistry, 2007. 58: p. 35-55.

5. Tian, T. and K. Burrage, *Stochastic models for regulatory networks of the genetic toggle switch.* Proc Natl Acad Sci U S A, 2006. 103(22): p. 8372-7.

6. Gomez-Uribe, C.A. and G.C. Verghese, *Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations.* The Journal of chemical physics, 2007. 126(2): p. 024109-12.

7. Lee, C.H., K.-H. Kim, and P. Kim, *A moment closure method for stochastic reaction networks.* The Journal of chemical physics, 2009. 130(13): p. 134107-15.

8. Singh, A. and J. Hespanha. *LogNormal Moment Closures for Biochemical Reactions.* in *In Proc. of the 45th Conf. on Decision and Contr.* 2006.

9. Milner, P., C.S. Gillespie, and D.J. Wilkinson, *Moment closure approximations for stochastic kinetic models with rational rate laws.* Mathematical Biosciences, 2011. 231(2): p. 99-104.

10. Chevalier, M.W. and H. El-Samad, *A rigorous framework for multiscale simulation of stochastic cellular networks.* The Journal of chemical physics, 2009. 131(5): p. 054102-17.

11. Voit, E.O., *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Vol. xii. 2000: Cambridge University Press.

12. Savageau, M.A., *Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions.* Journal of Theorectical Biology, 1969 a. 25(3): p. 365-9.

13. Savageau, M.A., *Biochemical systems analysis. A study of function and design in molecular biology*. Vol. xvii. 1976: Addison-Wesley.

14. Savageau, M., *Michaelis-Menten mechanism reconsidered: implications of fractal kinetics.* Journal of Theoretical Biology, 1995. 176(1): p. 115-124.

15. Savageau, M.A., *Influence of fractal kinetics on molecular recognition.* Journal of Molecular Recognition, 1993. 6(4): p. 149-157.

16. Bajzer, Z., Huzak, M., Neff, K.L., Prendergast, F.G.., *Mathematical analysis of models for reaction kinetics in intracellular environments.* Mathematical Biosciences, 2008. 215(1): p. 35-47.

17. Neff, K.L., *Biochemical reaction kinetics in dilute and crowded solutions: Predictions of macroscopic and mesoscopic models and experimental observations*. 2010, Mayo Clinic: Rochester, MN.

18. Neff, Kevin L., Offord, Chetan P, Caride, Ariel J, Strehler, Emanuel E, Prendergast, Franklyn G, Bajzer, eljko , *Validation of Fractal-Like Kinetic Models by Time-Resolved Binding Kinetics of Dansylamide and Carbonic Anhydrase in Crowded Media.* Biophysical journal, 2011. 100(10): p. 2495-2503.

19. Chou, I.-C. and E.O. Voit, *Recent developments in parameter estimation and structure identification of biochemical and genomic systems.* Math. Biosc. , 2009. 219: p. 57-83.

20. Walton, R.J., Preston, C. J., Bartlett, M., Smith, R.,Russell, R. G. G., *Biochemical measurements in Paget's disease of bone.* European Journal of Clinical Investigation, 1977. 7(1): p. 37-39.

21. Koch, A.L., *The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly.* Journal of Theoretical Biology, 1966. 12(2): p. 276-290.

22. Limpert, E., W.A. Stahel, and M. Abbt, *Log-normal Distributions across the Sciences: Keys and Clues.* BioScience, 2001. 51(5): p. 341.

23. Law, A.M. and W.D. Kelton, *Simulation Modeling and Analysis*. 3 ed. 2000, Boston: Mc.Graw Hill.

24. Gillespie, D. and L. Petzold, *Numerical Simulation for Biochemical Kinetics*, in *In System Modelling in Cellular Biology*, Z. Szallasi, J. Stelling, and V. Periwal, Editors. 2006, MIT Press.

25. Wolkenhauer, O., Ullah M., Kolch W., Cho K., , *Modelling and Simulation of IntraCellular Dynamics: Choosing an Appropriate Framework* IEEE Transactions on NanoBioscience, 2004. 3: p. 200-207.

26. Paulsson, J., O.G. Berg, and M. Ehrenberg, *Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation.* Proceedings of the National Academy of Sciences, 2000. 97(13): p. 7148-7153.

27. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators.* Nature, 2000. 403(6767): p. 335-338.

28. Bennett, M.R., Volfson, D., Tsimring, L., Hasty, J. , *Transient Dynamics of Genetic Regulatory Networks.* Biophysical journal, 2007. 92(10): p. 3501-3512.

29. Segel, L.A., *On the validity of the steady state assumption of enzyme kinetics.* Bull. Math. Biol. , 1988. 50: p. 579-593.

30. Michaelis, L. and M.L. Menten, *Die Kinetik der Invertinwirkung.* Biochem. Zeitschrift 1913. 49: p. 333-369.

31. Savageau, M.A. and E.O. Voit, *Recasting Nonlinear Differential-Equations As S-Systems - A Canonical Nonlinear Form.* Mathematical Bioscience, 1987. 87: p. 83-115.

32. Twomey, A., *On the Stochastic Modelling of Reaction-Diffusion Processes*. 2007, University of Oxford.

Figures

Figure 1. Scheme of reversible reaction with feedback controls

S_3 inhibits the forward reaction and S_1 activates the reverse reaction,

Figure 2. Comparative simulation results for a reversible reaction with feedback controls

In all panels, the *x*-axis denotes time in seconds and the *y*-axis represents the number of molecules of species S_1. The upper and lower panels use two different sets of initial numbers of molecules, namely:

$\left(x_1(0), x_2(0), x_3(0), U\right) = \left(5, 5, 6, 1\mu m^3\right)$ and $\left(x_1(0), x_2(0), x_3(0), U\right) = \left(100, 100, 120, 20\mu m^3\right)$, respectively. Other simulation parameters are

$\left(f_1, f_2, f_3, g_1, g_3, k_f, k_g\right) = \left(1.3, 1.8, -1, 1, 1, 0.5, 0.5\right)$. In both the upper and lower panels, the first column compares the time evolution of S_1 molecules by different methods: the black

line shows the ODE solution of Equation (52) for x_1; the blue lines are the solutions of Equation (53) for μ_1 and for $\mu_1 \pm \sigma_1$, respectively. The red dotted lines framing the mean indicate the 95% confidence interval.

The second column shows the propensity adjustment functions for the forward reaction (solid line) and the backward reaction (dashed line). The third column shows 100 independent stochastic simulations with propensity adjustment (blue means and error bars), in comparison with the ODE (Equation (52)) prediction (black line). The fourth column shows a second set of 100 independent stochastic simulations without propensity adjustment (blue means and error bars), in comparison with the ODE (Equation (52)) prediction (black line). The red dotted lines framing the mean in columns 3 and 4 again indicate the 95% confidence intervals.

Figure 3. Reaction scheme of the Repressilator

Gene G_1 codes for protein x_1, whose dimer y_1 represses the transcription of gene G_2. Similarly, y_2, the dimer of gene G_2's protein product x_2, represses the transcription of gene G_3, and y_3, the dimer of gene G_3's protein product x_3, represses the transcription of gene G_1.

Figure 4. Scaling of the Repressilator equations changes the oscillation period in the stochastic simulation

Solid lines represent solutions of ODEs (56), while dotted lines are trajectories of a stochastic simulation; blue lines represent x_1 and black lines represent m_1.

Figure 5. Power-law approximation of $p(x_i)^{-1}$

Left panel: Approximation of $y_i = p(x_i)^{-1}$ by a straight line in log-log space. Right panel: Corresponding power-law function in Cartesian space. Both axes are unitless.

Figure 6. Comparison of the dynamics of the Repressilator models using the original ODEs (55), the GMA approximation, and the moment approach based on the GMA approximation. The mean of the moment approach based on the GMA approximation fits the original ODEs (55) very well up to about *t*=400s. Black bold line: solution of the original ODEs (55); black dashed line: the GMA approximation; Red line: mean of the moment approach based on the GMA approximation; red dashed lines framing around the red line: mean $\pm$ standard deviation, which were produced with the moment approach. *x*-axis is time in second, *y*-axis is the number of x_1 molecules (unitless).

Figure 7. Stochastic focusing

The first panel from the top compares the time evolution of product molecules *P* obtained with different methods: the black line represents the solution of ODE model (64) for *P*; the blue solid line and blue dashed lines are the solutions of the moment-based model (65) for μ_P and $\mu_P \pm \sigma_P$, respectively. The second panel indicates that the stochastic simulations without propensity adjustment (blue error bar) diverge from the prediction by the ODE model (64) (black line). The third panel shows the propensity adjustment function $paf_4(t) = \exp\left(-f_I\, f_S \operatorname{cov}\left[\log I(t), \log S(t)\right]\right)$ for the reaction $R_4 : I + S \xrightarrow{k_4} S$. The bottom panel demonstrates that the propensity adjustment function paf_4 achieves convergence between the stochastic simulation and the ODE model (64) (black line): the blue error bars

were computed from 100 independent stochastic simulations with propensity adjustment paf_4. The simulation parameters are

$(i(0), p(0), s(0), k_1, k_2, k_3, k_4, k_5, k_6, f_I, f_S) = (0,10,100,10^4,10^3,1,9.9\times10^3,10^4,10^3,1.1,0.9)$;

at t=0.1, the value of k_5 changes from 10^4 to 0.5×10^4.

Additional files

Additional file 1 –Derivation of the mean and variance of a power-law function of random variables

Additional file 2 –Computation of approximate mean and covariance for a generic propensity function to be used in stochastic simulations

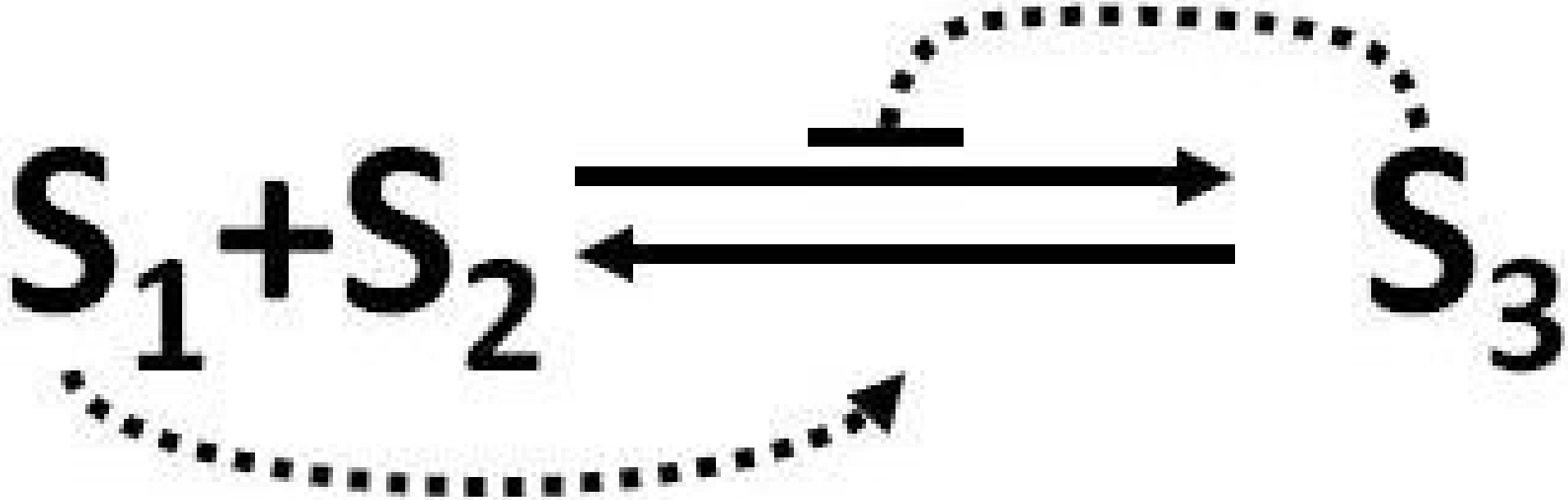


Figure 1

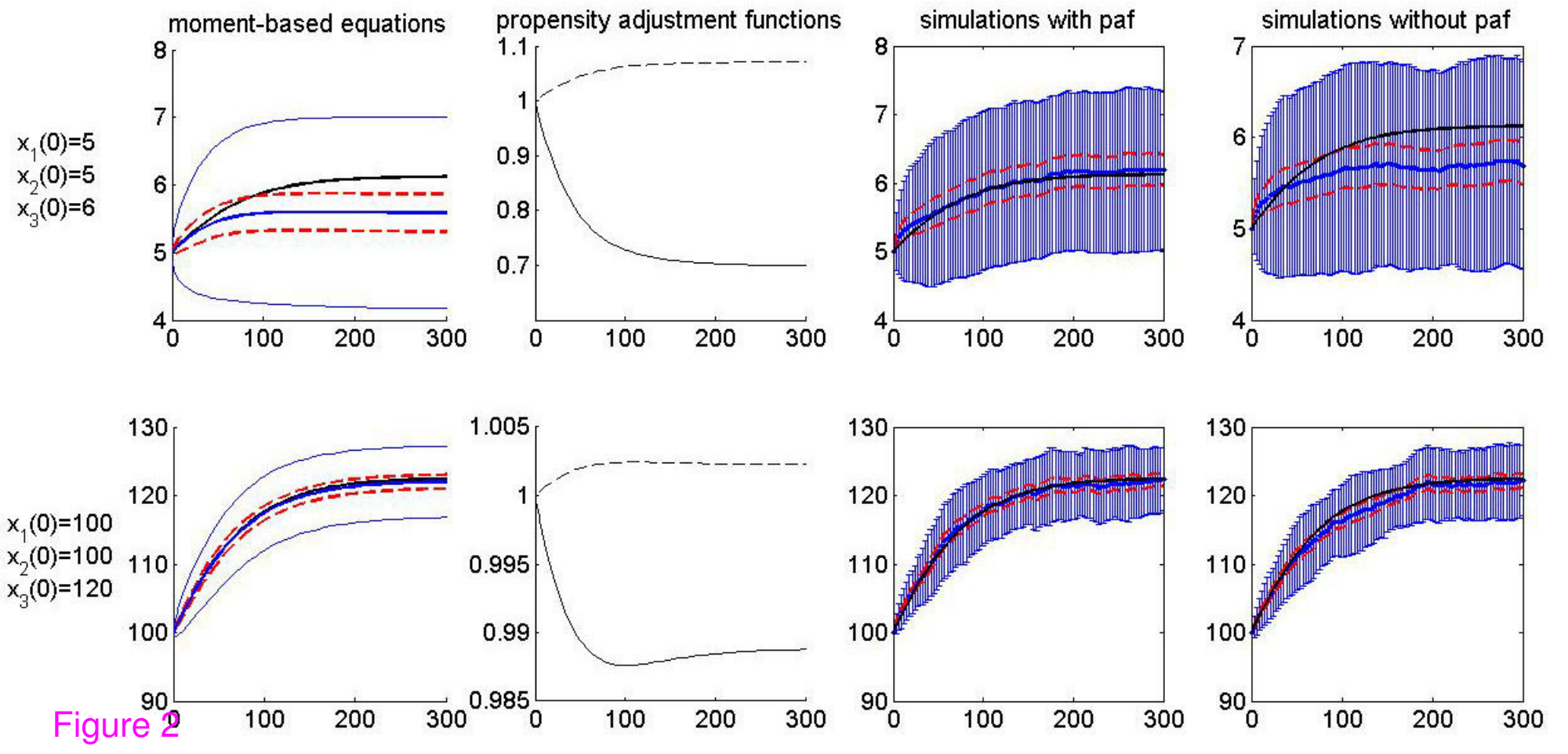

moment-based equations
propensity adjustment functions
simulations with paf
simulations without paf
x1(0)=5
x2(0)=5
x3(0)=6
x1(0)=100
x2(0)=100
x3(0)=120

Figure 2

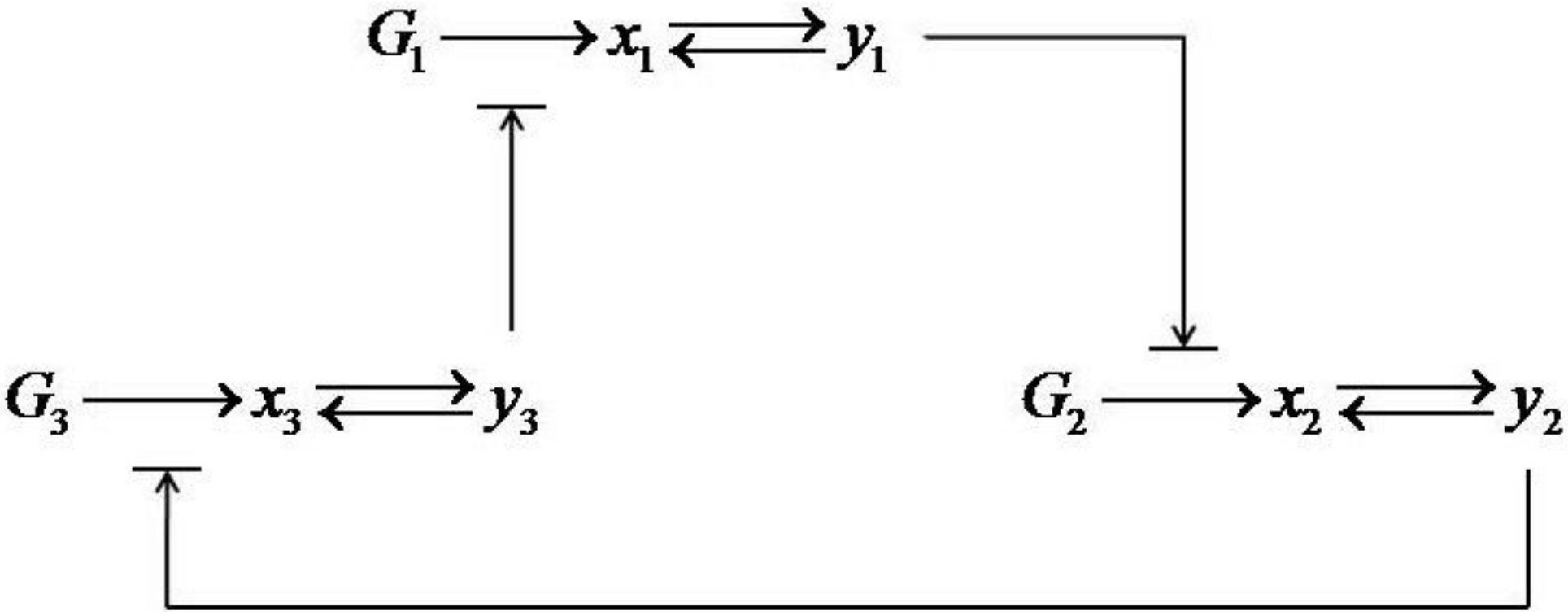


Figure 3

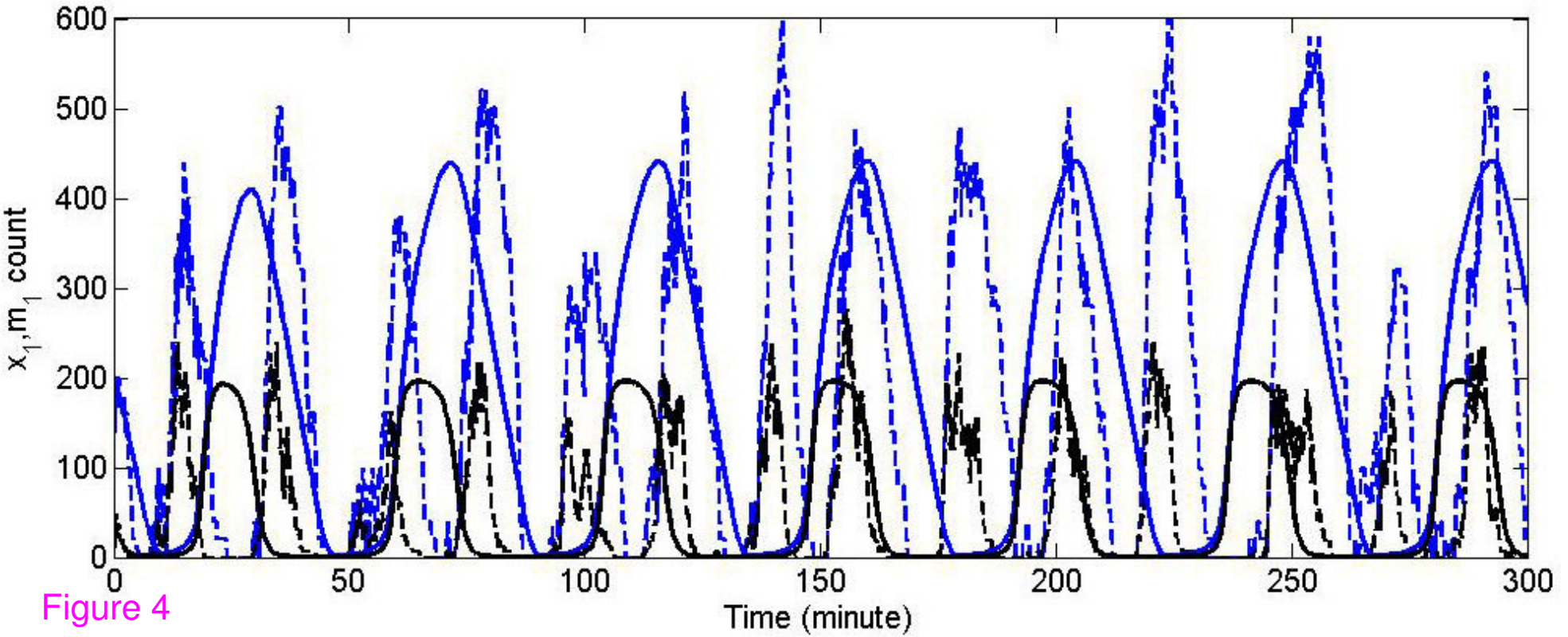
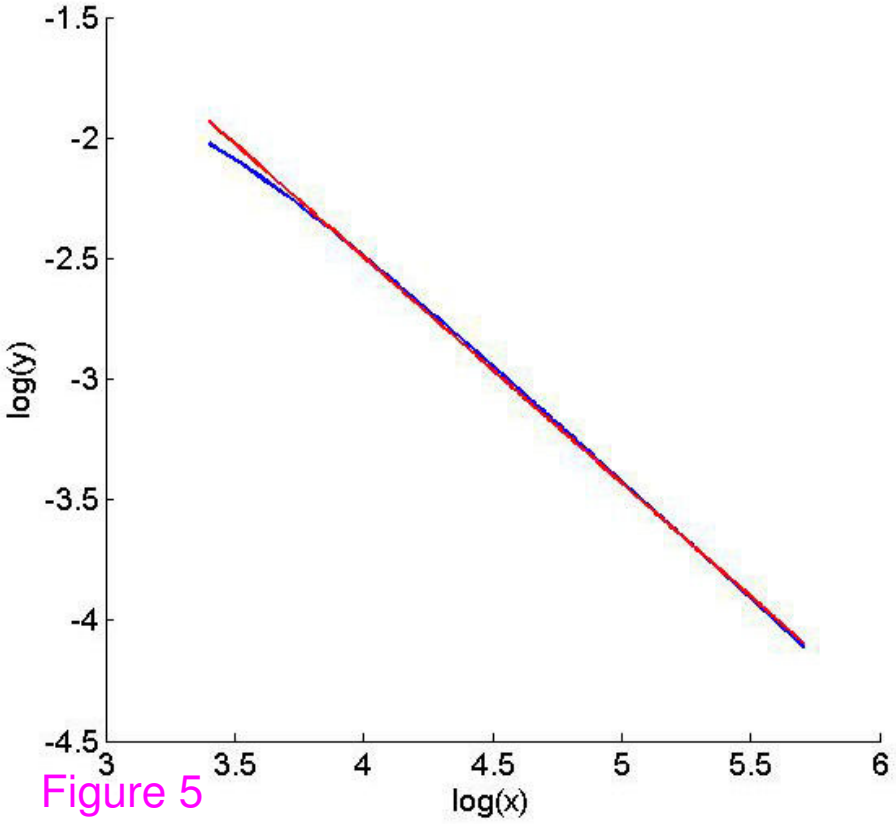

600
500
400
300
200
100
0
x_1, m_1 count
0
50
100
150
200
250
300
Time (minute)

Figure 4


-1.5
-2
-2.5
-3
-3.5
-4
-4.5
log(y)
3
3.5
4
4.5
5
5.5
6
log(x)

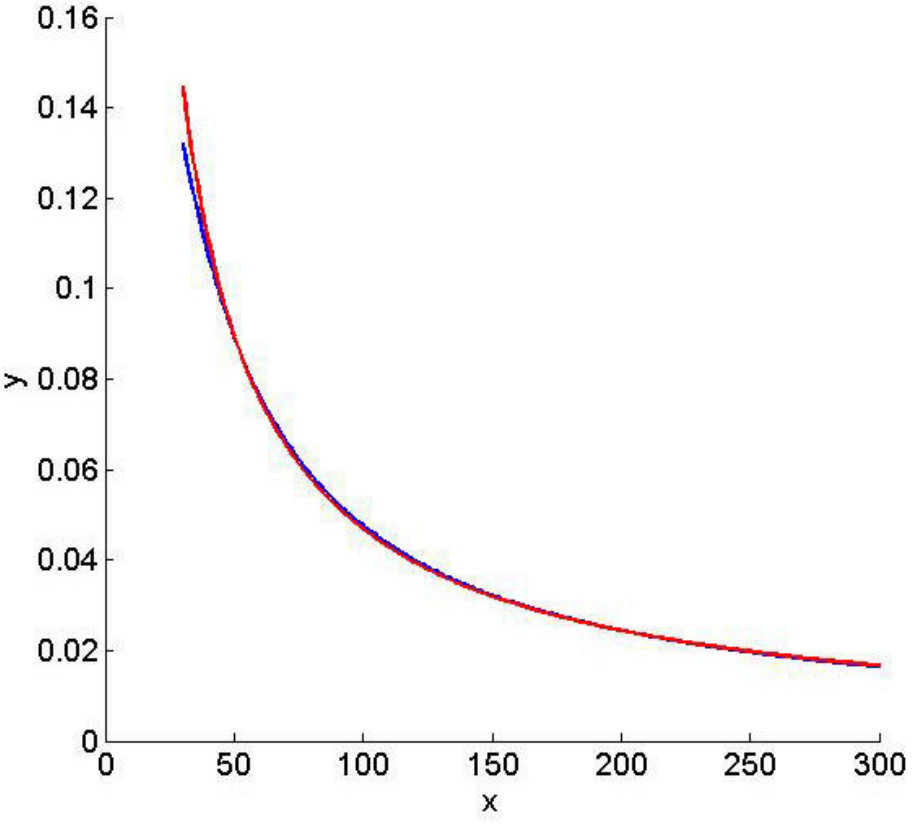

0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
y
0
50
100
150
200
250
300
x

Figure 5

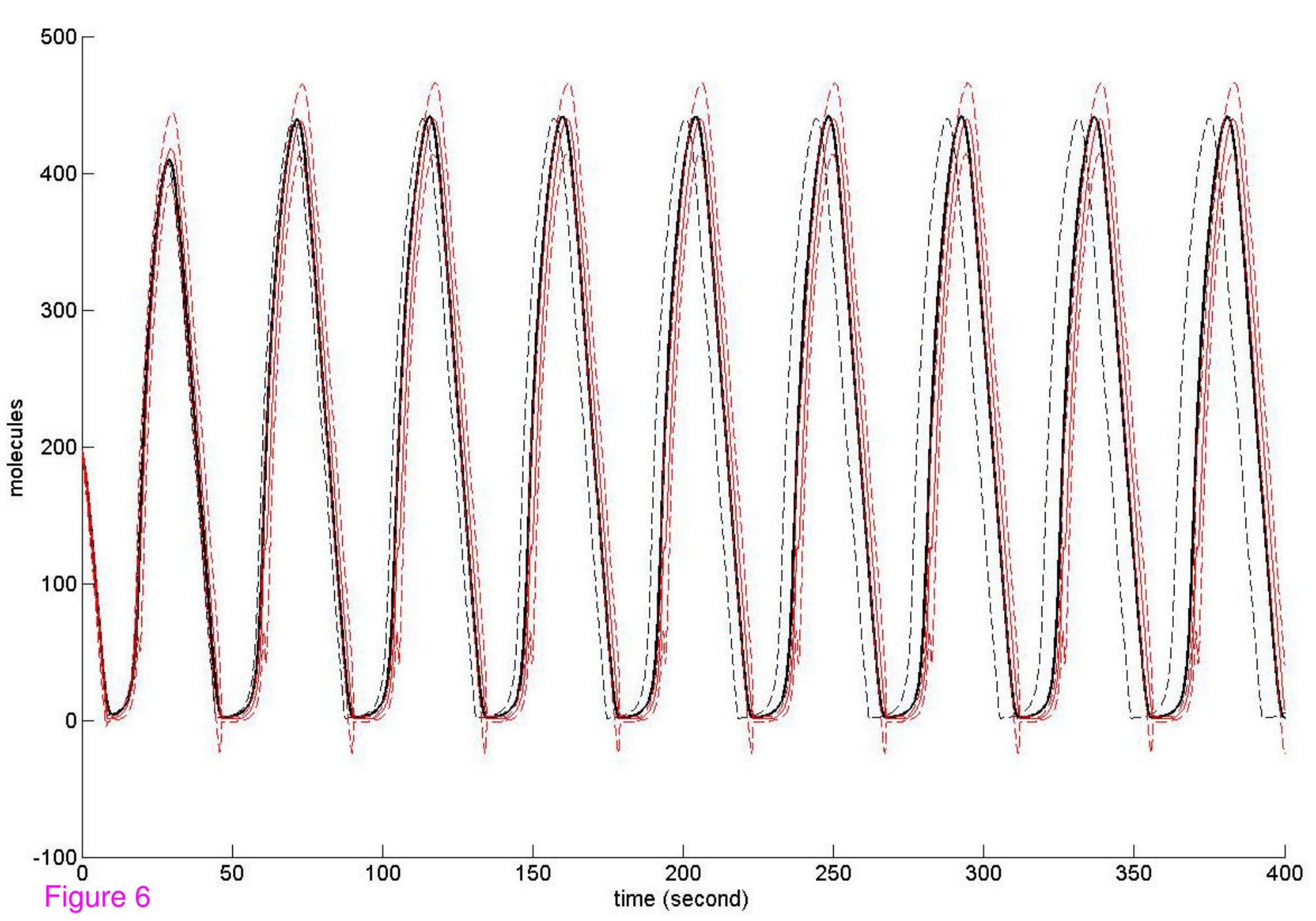

500
400
300
200
100
0
-100
molecules
0
50
100
150
200
250
300
350
400
time (second)

Figure 6

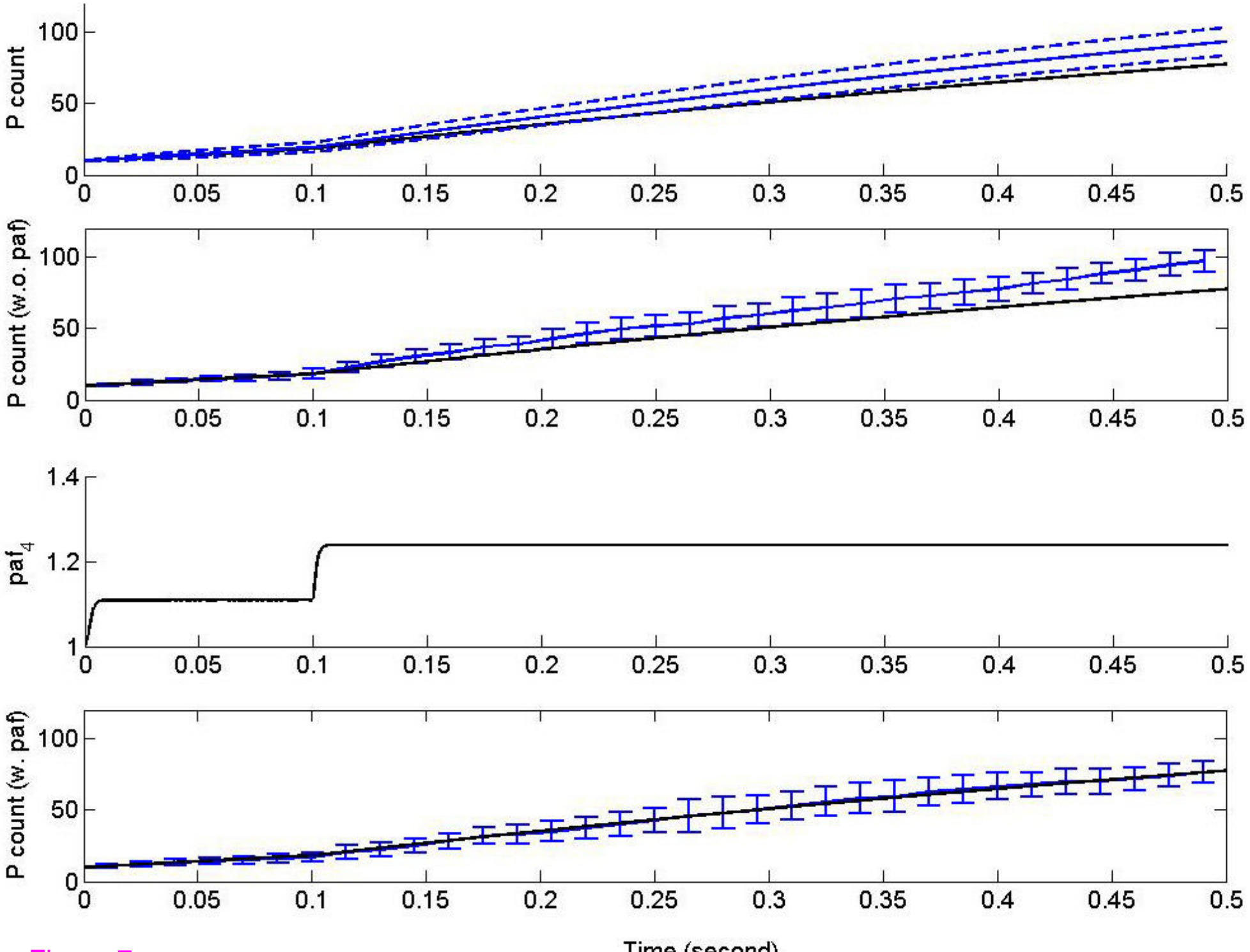

P count
100
50
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
P count (w.o. paf)
100
50
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
paf4
1.4
1.2
1
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
P count (w. paf)
100
50
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5
Time (second)

Figure 7

Additional files provided with this submission:

Additional file 1: Additional file 1.pdf, 122K
http://www.biomedcentral.com/imedia/2067161659573436/supp1.pdf
Additional file 2: Additional file 2.pdf, 132K
http://www.biomedcentral.com/imedia/4628526225734363/supp2.pdf